(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10363 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US01/23716

(22) International Filing Date: 26 July 2001 (26.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/221,679	28 July 2000 (28.07.2000)	US
60/223,272	3 August 2000 (03.08.2000)	US
60/224,309	10 August 2000 (10.08.2000)	US
60/226,728	18 August 2000 (18.08.2000)	US
60/229,254	30 August 2000 (30.08.2000)	US
60/231,366	8 September 2000 (08.09.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). WANG, Yumei, E. [CN/US]; 4624 Strawberry Park Drive, San Jose, CA 95129 (US). STEWART, Elisabeth, A. [US/US]; 1767 Monticello Road, San Mateo, CA 94402 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054

(US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). WARREN, Bridget, A. [US/US]; 10130 Parwood Drive #2, Cupertino, CA 95014 (US). NGUYEN, Danniel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN PHOSPHATASES

(57) Abstract: The invention provides human protein phosphatases (PP) and polynucleotides which identify and encode PP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PP.

PROTEIN PHOSPHATASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of protein phosphatases and

5   to the use of these sequences in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of protein phosphatases.

BACKGROUND OF THE INVENTION

10   Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues

15   serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, but can also occur along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak

20   signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

25   Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle

30   and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle

35   contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic

metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE

5   PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A

10   number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Opin.

15   Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit

20   additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE

25   PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through

30   the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases,

35   cyclin-dependent kinases, and the IκB kinases (reviewed in Millward et al., supra). PP2A is itself a

2

substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the

5   phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, supra).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the

10   calcium/calmodulin complex. Calcineurin is the target of the immunosuppresant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates

15   functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, supra).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide

20   repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids in vitro and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent

25   cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

30   In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site

35   consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate

3

moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently

5   affects the substrate specificity of the first.. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte

10   activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, supra). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2

15   domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, supra). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990)

20   Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The

25   MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266. Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine

30   phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, supra).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are

35   primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C.

The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, supra). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and

5   apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are

10   associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each

15   centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et

20   al. (1989) Eur. J. Biochem. 183:611-616). LAP-deficient mice have progressive skeletal disorder and an increased disposition toward generalized seizures (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) *200950).

25   PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A.G. et al. (1977) Cancer Res. 37:4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated

30   homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and M.F. Lin (1998) J. Biol. Chem. 34:22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-

35   coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein,

EPS15, which is mediated by the C-terminal region of synatojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be

5   involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling

10   (Cremona, O. et al. (1999) Cell 99:179-188).

The discovery of new protein phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the

15   expression of nucleic acid and amino acid sequences of protein phosphatases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, protein phosphatases, referred to collectively as "PP" and individually as "PP-1," "PP-2," "PP-3," "PP-4," "PP-5," "PP-6," "PP-7," "PP-8," "PP-9,"

20   and "PP-10." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

25   consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the

30   group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

35   In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of

6

SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group

5   consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino

10   acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

15   of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a)

20   culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

25   sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group

30   consisting of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID

35   NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide

complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group

5   consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The

10   method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if

15   present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of

20   SEQ ID NO:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain

25   reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a

30   naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the

35   composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-

10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an

5   agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an

10   immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of

15   treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide

20   comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) exposing a sample comprising the

25   polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment the composition.

30   The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test

5   compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected

10   from the group consisting of SEQ ID NO:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the

15   activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in

20   altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said

25   method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a

30   polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a

polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary

5   to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the

10   treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

15   Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for

20   analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

25   Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

30

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

35   particular embodiments only, and is not intended to limit the scope of the present invention which

will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

5   reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be

10   used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

15   **DEFINITIONS**

"PP" refers to the amino acid sequences of substantially purified PP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of

20   PP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

An "allelic variant" is an alternative form of the gene encoding PP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in

25   polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

30   "Altered" nucleic acid sequences encoding PP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PP or a polypeptide with at least one functional characteristic of PP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PP, and improper or unexpected hybridization to allelic variants, with

35   a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PP. The

encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as

5 the biological or immunological activity of PP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and

10 valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid

15 sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity

20 of PP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PP either by directly interacting with PP or by acting on components of the biological pathway in which PP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments

25 thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly

30 used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies

35 which bind specifically to antigenic determinants (particular regions or three-dimensional structures

13

on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA;
5   RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once
10   introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical
15   functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid
20   sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.
25   Compositions comprising polynucleotide sequences encoding PP or fragments of PP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).
30   "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison
35   WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and

14

assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows

5   amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,

30   (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

35   Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

40   A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

5   "Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of PP or the polynucleotide encoding PP which is identical

10   in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid

15   residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

20   embodiments.

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish

25   SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies

30   SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

35   A "full length" polynucleotide sequence is one containing at least a translation initiation

16

codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

5   The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

10   Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS

15   8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms

20   is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with

25   other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to

30   compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

35   *Open Gap: 5 and Extension Gap: 2 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

5   Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length

10   supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid

15   sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative

20   substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of

25   polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise

30   comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

35   *Gap x drop-off: 50*

18

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence,
for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for
example, over the length of a fragment taken from a larger, defined polypeptide sequence, for
instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least
150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment
length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be
used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain
DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for
chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid
sequence in the non-antigen binding regions has been altered so that the antibody more closely
resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a
complementary strand through base pairing under defined hybridization conditions. Specific
hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.
Specific hybridization complexes form under permissive annealing conditions and remain hybridized
after the "washing" step(s). The washing step(s) is particularly important in determining the
stringency of the hybridization process, with more stringent conditions allowing less non-specific
binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive
conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill
in the art and may be consistent among hybridization experiments, whereas wash conditions may be
varied among experiments to achieve the desired stringency, and therefore hybridization specificity.
Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about
1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature
under which the wash step is carried out. Such wash temperatures are typically selected to be about
5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic
strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of
the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and
conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al.
(1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press,

Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC

5   concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions

10 will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A

15 hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

20 sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

25   An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides,

30 polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological,

35 functional, or immunological properties of PP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

5   "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

10   "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

15   "Post-translational modification" of an PP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PP.

       "Probe" refers to nucleic acid sequences encoding PP, their complements, or fragments
20   thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target
25   DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

       Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,
30   or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

       Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold
35   Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular

Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR
Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs
can be derived from a known sequence, for example, by using computer programs intended for that
purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge
5   MA).

Oligonucleotides for use as primers are selected using software known in the art for such
purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to
100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to
5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer
10   selection programs have incorporated additional features for expanded capabilities. For example, the
PrimOU primer selection program (available to the public from the Genome Center at University of
Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from
megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3
primer selection program (available to the public from the Whitehead Institute/MIT Center for
15   Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which
sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the
selection of oligonucleotides for microarrays. (The source code for the latter two primer selection
programs may also be obtained from their respective sources and modified to meet the user's specific
needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping
20   Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments,
thereby allowing selection of primers that hybridize to either the most conserved or least conserved
regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both
unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and
polynucleotide fragments identified by any of the above selection methods are useful in hybridization
25   technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to
identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of
oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence
that is made by an artificial combination of two or more otherwise separated segments of sequence.
30   This artificial combination is often accomplished by chemical synthesis or, more commonly, by the
artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques
such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have
been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a
recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter
35   sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to

transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

5   A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid,
10   amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the
15   nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PP, nucleic acids encoding PP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in
20   solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding
25   molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free,
30   preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,
35   chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,

microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

5   "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or

10   viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to

15   animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in

20   vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention

25   into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-

30   1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have

35   significant identity to a reference molecule, but will generally have a greater or lesser number of

24

polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding
polypeptide may possess additional functional domains or lack domains that are present in the
reference molecule. Species variants are polynucleotide sequences that vary from one species to
another. The resulting polypeptides will generally have significant amino acid identity relative to

5   each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene
between individuals of a given species. Polymorphic variants also may encompass "single nucleotide
polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The
presence of SNPs may be indicative of, for example, a certain population, a disease state, or a
propensity for a disease state.

10   A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having
at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of
the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-
1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at
least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least

15   94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence
identity over a certain defined length of one of the polypeptides.

THE INVENTION

          The invention is based on the discovery of new human protein phosphatases (PP), the
20   polynucleotides encoding PP, and the use of these compositions for the diagnosis, treatment, or
prevention of immune system disorders, neurological disorders, developmental disorders, and cell
proliferative disorders.

          Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide
sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a
25   single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted
by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte
polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and
an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

30   Table 2 shows sequences with homology to the polypeptides of the invention as identified by
BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the
polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte
polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3
shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog.
35   Column 4 shows the probability score for the match between each polypeptide and its GenBank

homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

     Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding
5   Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and
10   motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

     Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are protein phosphatases. For example, SEQ ID NO:2 is 98% identical to mouse putative protein phosphatase type 2C (GenBank ID g4325051) as
15   determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.0e-89, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a protein phosphatase 2C domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS,
20   MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:2 is a protein phosphatase 2C. In an alternative example, SEQ ID NO:4 is 46% identical to human protein phosphatase (GenBank ID g6692782) as determined by BLAST. (See Table 2.) The BLAST probability score is 2.0e-33. SEQ ID NO:4 also contains a dual specificity phosphatase, catalytic domain as determined by searching for statistically significant matches in the HMM-based PFAM
25   database. (See Table 3.) Data from BLAST-DOMO analysis provides further corroborative evidence that SEQ ID NO:4 is a dual specificity protein phosphatase. In an alternative example, SEQ ID NO:6 is 45% identical to murine lysosomal acid phosphatase (GenBank ID g52871) as determined by BLAST. (See Table 2.) The BLAST probability score is 2.3e-83. SEQ ID NO:6 also contains a histidine acid phosphatase domain as determined by searching for statistically significant matches in
30   the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is an acid phosphatase. In an alternative example, SEQ ID NO:7 is 52% identical to mouse neuronal tyrosine threonine phosphatase 1 (GenBank ID g1781037) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.3e-131. SEQ ID NO:7 also contains a dual specificity phosphatase active site
35   domain as determined by searching for statistically significant matches in the HMM-based PFAM

database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a tyrosine phosphatase. In an alternative example, SEQ ID NO:8 is 61% identical to human tyrosine phosphatase (GenBank ID g6650693) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.0e-89. SEQ ID NO:8 also

5   contains a transmembrane domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from MOTIFS analyses provide further corroborative evidence that SEQ ID NO:8 is tyrosine specific protein phosphatase. Tyrosine phosphatases are one of two general categories of protein phosphatases. In an alternative example, SEQ ID NO:10 is 55% identical to human mitogen-activated protein kinase phosphatase (GenBank ID

10   g9294745) as determined by BLAST. (See Table 2.) The BLAST probability score is 1.3e-50. SEQ ID NO:10 also contains a dual specificity phosphatase catalytic domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a protein kinase phosphatase. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID

15   NO:9 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence

20   identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and

25   related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective

30   full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6024861H1 is the identification number of an Incyte cDNA sequence, and TESTNOT11 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from

35   pooled cDNA libraries (e.g., 71907683V1). Alternatively, the identification numbers in column 5

27

may refer to GenBank cDNAs or ESTs (e.g., g2114900) which contributed to the assembly of the full

length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those

5   sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon

10   stitching" algorithm. For example, FL_$XXXXXX$_N_1_N_2_$YYYYY$_N_3_N_4 represents a "stitched" sequence in which $XXXXXX$ is the identification number of the cluster of sequences to which the algorithm was applied, and $YYYYY$ is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of

15   exons brought together by an "exon-stretching" algorithm. For example, FL$XXXXXX$_g$AAAAA$_g$BBBBB$_1_N is the identification number of a "stretched" sequence, with $XXXXXX$ being the Incyte project identification number, g$AAAAA$ being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g$BBBBB$ being the GenBank identification number or NCBI RefSeq identification number of the

20   nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g$BBBBB$).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from

25   genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

28

INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant

5   Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors

10   which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PP variants. A preferred PP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PP amino acid sequence, and which contains at least one functional or structural characteristic of PP.

15   The invention also encompasses polynucleotides which encode PP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes PP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone

20   is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PP. A particular aspect of the invention encompasses a variant of a

25   polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PP.

30   It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could

be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PP and its variants are generally capable of
5   hybridizing to the nucleotide sequence of the naturally occurring PP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular
10 codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PP and PP
15 derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of
20 hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:11-20 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

25 Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE
30 amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system
35 (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences

are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PP may be extended utilizing a partial nucleotide
5   sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown
10   sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme
15   digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries
20   and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

25   When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

30   Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate
35   software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire

process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

5   In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PP may be cloned in recombinant DNA molecules that direct expression of PP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PP.

       The nucleotide sequences of the present invention can be engineered using methods generally
10   known in the art in order to alter PP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction
15   sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

       The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or
20   improve the biological properties of PP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of
25   DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby
30   maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

       In another embodiment, sequences encoding PP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)
35   Alternatively, PP itself or a fragment thereof may be synthesized using chemical methods. For

example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the

5   amino acid sequence of PP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or . a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

10   The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PP, the nucleotide sequences encoding PP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in

15   a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where

20   sequences encoding PP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both

25   natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PP and appropriate transcriptional and translational control

30   elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences

35   encoding PP. These include, but are not limited to, microorganisms such as bacteria transformed with

recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with
yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus);
plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV,
or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or

5   animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster
(1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA
91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO
J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New
York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and

10   Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses,
adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for
delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di
Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci.
USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al.

15   (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)
The invention is not limited by the host cell employed.

        In bacterial systems, a number of cloning and expression vectors may be selected depending
upon the use intended for polynucleotide sequences encoding PP. For example, routine cloning,
subcloning, and propagation of polynucleotide sequences encoding PP can be achieved using a

20   multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1
plasmid (Life Technologies). Ligation of sequences encoding PP into the vector's multiple cloning
site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of
transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for
in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of

25   nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol.
Chem. 264:5503-5509.) When large quantities of PP are needed, e.g. for the production of
antibodies, vectors which direct high level expression of PP may be used. For example, vectors
containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

        Yeast expression systems may be used for production of PP. A number of vectors containing

30   constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may
be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct
either the secretion or intracellular retention of expressed proteins and enable integration of foreign
sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A.
et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-

35   184.)

Plant systems may also be used for expression of PP. Transcription of sequences encoding
PP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in
combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.
6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock

5   promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al.
(1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.)
These constructs can be introduced into plant cells by direct DNA transformation or
pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology
(1992) McGraw Hill, New York NY, pp. 191-196.)

10   In mammalian cells, a number of viral-based expression systems may be utilized. In cases
where an adenovirus is used as an expression vector, sequences encoding PP may be ligated into an
adenovirus transcription/translation complex consisting of the late promoter and tripartite leader
sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain
infective virus which expresses PP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl.

15   Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus
(RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based
vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are

20   constructed and delivered via conventional delivery methods (liposomes, polycationic amino
polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet.
15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression
of PP in cell lines is preferred. For example, sequences encoding PP can be transformed into cell

25   lines using expression vectors which may contain viral origins of replication and/or endogenous
expression elements and a selectable marker gene on the same or on a separate vector. Following the
introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media
before being switched to selective media. The purpose of the selectable marker is to confer resistance
to a selective agent, and its presence allows growth and recovery of cells which successfully express

30   the introduced sequences. Resistant clones of stably transformed cells may be propagated using
tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These
include, but are not limited to, the herpes simplex virus thymidine kinase and adenine
phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et

35   al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic,

or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981)

5   J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to

10   quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PP is inserted within a marker gene sequence, transformed cells containing

15   sequences encoding PP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PP and that express PP

20   may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PP using either

25   specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g.,

30   Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and

35   may be used in various nucleic acid and amino acid assays. Means for producing labeled

36

hybridization or PCR probes for detecting sequences related to polynucleotides encoding PP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available,

5   and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic

10   agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors

15   containing polynucleotides which encode PP may be designed to contain signal sequences which direct secretion of PP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

20   phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct

25   modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the

30   screening of peptide libraries for inhibitors of PP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion

35   proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate

resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of

fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically

recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic

cleavage site located between the PP encoding sequence and the heterologous protein sequence, so

5   that PP may be cleaved away from the heterologous moiety following purification. Methods for

fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety

of commercially available kits may also be used to facilitate expression and purification of fusion

proteins.   \

In a further embodiment of the invention, synthesis of radiolabeled PP may be achieved in

10   vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These

systems couple transcription and translation of protein-coding sequences operably associated with the

T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid

precursor, for example, ^{35}S-methionine.

PP of the present invention or fragments thereof may be used to screen for compounds that

15   specifically bind to PP. At least one and up to a plurality of test compounds may be screened for

specific binding to PP. Examples of test compounds include antibodies, oligonucleotides, proteins

(e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of

PP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a

20   natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):

Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PP binds,

or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound

can be rationally designed using known techniques. In one embodiment, screening for these

compounds involves producing appropriate cells which express PP, either as a secreted protein or on

25   the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells

expressing PP or cell membrane fractions which contain PP are then contacted with a test compound

and binding, stimulation, or inhibition of activity of either PP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example,

30   the assay may comprise the steps of combining at least one test compound with PP, either in solution

or affixed to a solid support, and detecting the binding of PP to the compound. Alternatively, the

assay may detect or measure binding of a test compound in the presence of a labeled competitor.

Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural

product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PP activity, wherein PP is combined with at least one test compound, and the activity of PP in the presence of a

5   test compound is compared with the activity of PP in the absence of the test compound. A change in the activity of PP in the presence of the test compound is indicative of a compound that modulates the activity of PP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PP under conditions suitable for PP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PP may do so indirectly and need not come

10   in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of

15   human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host

20   genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred

25   to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell

30   lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a

35   polynucleotide encoding PP is injected into animal ES cells, and the injected sequence integrates into

the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PP, e.g., by secreting PP in its milk, may also serve as a convenient

5   source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PP and protein phosphatases. In addition, the expression of PP is closely associated with thalamus, pancreas, testis, brain, vascular, and fetal lung tissues, as well as colon

10 tissue pseudopolyps associated with multiple tubuvillous adenomas. Therefore, PP appears to play a role in immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders. In the treatment of disorders associated with increased PP expression or activity, it is desirable to decrease the expression or activity of PP. In the treatment of disorders associated with decreased PP expression or activity, it is desirable to increase the expression or

15 activity of PP.

Therefore, in one embodiment, PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common

20 variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing

25 spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

30 syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

35 complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal,

parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis

5 pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,

10 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and

15 toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's

20 syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea

25 and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

30 teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing PP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression

35 or activity of PP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those provided above.

5   In still another embodiment, an agonist which modulates the activity of PP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PP including, but not limited to, those listed above.

In a further embodiment, an antagonist of PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP. Examples of such disorders

10   include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds PP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PP.

In an additional embodiment, a vector expressing the complement of the polynucleotide

15   encoding PP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made

20   by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PP may be produced using methods which are generally known in the art. In

25   particular, purified PP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PP. Antibodies to PP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer

30   formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral

35   gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic

polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of

5   at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PP may be prepared using any technique which provides for the

10   production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

15   In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single

20   chain antibodies may be adapted, using methods known in the art, to produce PP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte

25   population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PP may also be generated. For example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin

30   digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired

35   specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PP epitopes is generally used, but a competitive binding assay may also be
5   employed (Pound, supra).

        Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a
10 determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PP epitopes, represents the average affinity, or avidity, of the antibodies for PP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PP-antibody
15 complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

20   The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and
25 guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

        In another embodiment of the invention, the polynucleotides encoding PP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,
30 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

        In therapeutic use, any gene delivery system suitable for introduction of the antisense
35 sequences into appropriate target cells can be used. Antisense sequences can be delivered

44

intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PP expression or regulation causes disease, the expression of PP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PP are treated by constructing mammalian expression vectors encoding PP and introducing these vectors by mechanical means into PP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics,

Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and

5   PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and

10   H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PP from a normal individual.

15   Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al.

20   (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PP under the control of an independent promoter or the retrovirus long

25   terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in

30   an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining

35   retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")

discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4[+] T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-

5   7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PP to cells which have one or more genetic abnormalities with respect to

10   the expression of PP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"),

15   hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PP to target cells which have one or more genetic abnormalities with

20   respect to the expression of PP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.

25   169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also

30   taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation

35   of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of

ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the

5   SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PP into the alphavirus genome in place

10 of the capsid-coding region results in the production of a large number of PP-coding RNAs and the synthesis of high levels of PP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al.

15 (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

20 Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using

25 triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of

30 RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by

35 scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,

GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary

5   oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA

10   sequences encoding PP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

15   modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine,

20   cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not

25   limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PP

30   expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PP may be therapeutically useful, and in the treatment of disorders associated with decreased PP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in

35   altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a

5   library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is

10   detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the

15   polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a

20   combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable

25   for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

30   Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable

35   excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins.

Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PP, antibodies to PP, and mimetics, agonists, antagonists, or inhibitors of PP.

The compositions utilized in this invention may be administered by any number of routes

5 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the

10 case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration

15 without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of

20 macromolecules comprising PP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999)

25 Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and

30 routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PP or fragments thereof, antibodies of PP, and agonists, antagonists or inhibitors of PP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the

35 ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of

the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of

5 circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the

10 active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

15 Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,

20 conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PP may be used for the diagnosis of disorders characterized by expression of PP, or in assays to monitor patients being treated with PP or agonists, antagonists, or inhibitors of PP. Antibodies useful for diagnostic purposes may be

25 prepared in the same manner as described above for therapeutics. Diagnostic assays for PP include methods which utilize the antibody and a label to detect PP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

30 A variety of protocols for measuring PP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PP expression. Normal or standard values for PP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PP under conditions suitable for complex formation. The amount of standard complex formation may be

35 quantitated by various methods, such as photometric means. Quantities of PP expressed in subject,

52

control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences,

5   complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PP, and to monitor regulation of PP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide

10   sequences, including genomic sequences, encoding PP or closely related molecules may be used to identify nucleic acid sequences which encode PP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PP, allelic variants, or related

15   sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:11-20 or from genomic sequences including promoters, enhancers, and introns of the PP gene.

20   Means for producing specific hybridization probes for DNAs encoding PP include the cloning of polynucleotide sequences encoding PP or PP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups,

25   for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PP may be used for the diagnosis of disorders associated with expression of PP. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of

30   Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies,

35   ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,

53

autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

5   syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

10 complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis

15 pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,

20 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and

25 toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's

30 syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea

35 and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

sensorineural hearing loss; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

5   teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding PP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-

10   like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PP may be labeled by standard methods and added to a fluid or tissue sample

15   from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate

20   the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a

25   fragment thereof, encoding PP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the

30   presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several

35   days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals

5   to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide

10   encoding PP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences

15   encoding PP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PP are used to amplify DNA using the

20   polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such

25   as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the

30   alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

35   et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be

accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the
5   polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor
10 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her
15 pharmacogenomic profile.

In another embodiment, PP, fragments of PP, or antibodies specific for PP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to
20 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by
25 hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

30 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of
35 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental

compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test

5   compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes

10   are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of

15   Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the

20   treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

25   Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under

30   given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson,

35   supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by

staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

density of each protein spot is generally proportional to the level of the protein in the sample. The

optical densities of equivalently positioned protein spots from different samples, for example, from

biological samples either treated or untreated with a test compound or therapeutic agent, are

5   compared to identify any changes in protein spot density related to the treatment. The proteins in the

spots are partially sequenced using, for example, standard methods employing chemical or enzymatic

cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by

comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the

polypeptide sequences of the present invention. In some cases, further sequence data may be

10   obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PP to quantify the

levels of PP expression. In one embodiment, the antibodies are used as elements on a microarray, and

protein expression levels are quantified by exposing the microarray to the sample and detecting the

levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111;

15   Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of

methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-

reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and

should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor

20   correlation between transcript and protein abundances for some proteins in some tissues (Anderson,

N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be

useful in the analysis of compounds which do not significantly affect the transcript image, but which

alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to

rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such

25   cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated

biological sample are separated so that the amount of each protein can be quantified. The amount of

each protein is compared to the amount of the corresponding protein in an untreated biological

30   sample. A difference in the amount of protein between the two samples is indicative of a toxic

response to the test compound in the treated sample. Individual proteins are identified by sequencing

the amino acid residues of the individual proteins and comparing these partial sequences to the

polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

35   sample containing proteins with the test compound. Proteins from the biological sample are

incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated

5   sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-

10 2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either

15 coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs),

20 yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the

25 inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man

30 (OMIM) World Wide Web site. Correlation between the location of the gene encoding PP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

35 Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23,

5   any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

        In another embodiment of the invention, PP, its catalytic or immunogenic fragments, or

10   oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PP and the agent being tested may be measured.

        Another technique for drug screening provides for high throughput screening of compounds

15   having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PP, or fragments thereof, and washed. Bound PP is then detected by methods well known in the art. Purified PP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,

20   non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

        In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PP specifically compete with a test compound for binding PP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PP.

25   In additional embodiments, the nucleotide sequences which encode PP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

        Without further elaboration, it is believed that one skilled in the art can, using the preceding

30   description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

        The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/221,679, U.S. Ser. No. 60/223,272, U.S. Ser. No. 60/224,309, U.S. Ser.

35   No. 60/226,728, U.S. Ser. No. 60/229,254, and U.S. Ser. No. 60/231,366, are expressly incorporated

by reference herein.

5   **EXAMPLES**

I.   Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database
(Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized
and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a
10   suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of
phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or
extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium
acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA
15   purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated
using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN,
Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was
isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA
purification kit (Ambion, Austin TX).

20   In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA
libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP
vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the
recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units
5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic
25   oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the
appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-
1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column
chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs
were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,
30   PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid
(Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto
CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells
including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX
DH10B from Life Technologies.

35   **II.   Isolation of cDNA Clones**

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid,

5 QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal

10 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III.   Sequencing and Analysis

15   Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared

20 using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI

25 protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing

30 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family

35 databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary

structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or

5   Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention

10   may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software

15   Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of

20   Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the

25   strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization

30   and amplification technologies are described in Table 4, column 4.

IV.   Identification and Editing of Coding Sequences from Genomic DNA

Putative protein phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a

35   variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and

S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan

5   predicted cDNA sequences encode protein phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for protein phosphatases. Potential protein phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as protein phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by

10   comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling

15   Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V.   Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

20   Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information,

25   generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated

30   but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent

35   type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared

.by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

5   Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

10   Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original 'GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences

15   were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

· **VI.   Chromosomal Mapping of PP Encoding Polynucleotides**

        The sequences which were used to assemble SEQ ID NO:11-20 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other

20   implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

25   had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

        Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

30   chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site

35   (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified

disease genes map within or in proximity to the intervals indicated above.

VII.   Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs

5   from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the

10   computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

15

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is

20   calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the

25   entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PP are analyzed with respect to the tissue

30   sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male;

35   germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas;

respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma,

5   cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PP Encoding Polynucleotides

10   Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30

15   nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

20   High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer

25   pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

30   The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

35   concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by

electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

5   sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site

10   overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following

15   parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing

20   primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

25   IX.   Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

30   software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based

35   hybridization analysis of human genomic DNA digested with one of the following endonucleases:

Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon
membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16
hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature
5   under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.
Hybridization patterns are visualized using autoradiography or an alternative imaging means and
compared.

X.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing
10   photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.),
mechanical microspotting technologies, and derivatives thereof. The substrate in each of the
aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999),
supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers.
Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link
15   elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding
procedures. A typical array may be produced using available methods and machines well known to
those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g.,
Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645;
Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

20   Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may
comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be
selected using software well known in the art such as LASERGENE software (DNASTAR). The
array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the
biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.
25   After hybridization, nonhybridized nucleotides from the biological sample are removed, and a
fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser
desorbtion and mass spectrometry may be used for detection of hybridization. The degree of
complementarity and the relative abundance of each polynucleotide which hybridizes to an element
on the microarray may be assessed. In one embodiment, microarray preparation and usage is
30   described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and
poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is
reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X
35   first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40

μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one

5   with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

10   then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification

15   uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

20   slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

25   Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

30   Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just

5   slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

10   **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide

15   containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,

20   Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

25   The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different

30   fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC

computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

5   emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

10   **XI.   Complementary Polynucleotides**

Sequences complementary to the PP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO

15   4.06 software (National Biosciences) and the coding sequence of PP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PP-encoding transcript.

XII.   Expression of PP

20   Expression and purification of PP is achieved using bacterial or virus-based expression systems. For expression of PP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

25   Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PP by

30   either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA

35   91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on

5   immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

10   (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PP obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX where applicable.

XIII.   Functional Assays

PP function is assessed by expressing the sequences encoding PP at physiologically elevated

15   levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

20   formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-

25   based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light

30   scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

35   The influence of PP on gene expression can be assessed using highly purified populations of

cells transfected with sequences encoding PP and either CD64 or CD64-GFP. CD64 and CD64-GFP
are expressed on the surface of transfected cells and bind to conserved regions of human
immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using
magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success

5   NY). mRNA can be purified from the cells using methods well known by those of skill in the art.
Expression of mRNA encoding PP and other genes of interest can be analyzed by northern analysis or
microarray techniques.

XIV.   Production of PP Specific Antibodies

PP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,

10   Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to
immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PP amino acid sequence is analyzed using LASERGENE software
(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is
synthesized and used to raise antibodies by means known to those of skill in the art. Methods for

15   selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well
described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A
peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-
Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

20   increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the
oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for
antipeptide and anti-PP activity by, for example, binding the peptide or PP to a substrate, blocking
with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-
rabbit IgG.

25 ### XV.   Purification of Naturally Occurring PP Using Specific Antibodies

Naturally occurring or recombinant PP is substantially purified by immunoaffinity
chromatography using antibodies specific for PP. An immunoaffinity column is constructed by
covalently coupling anti-PP antibody to an activated chromatographic resin, such as CNBr-activated
SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed

30   according to the manufacturer's instructions.

Media containing PP are passed over the immunoaffinity column, and the column is washed
under conditions that allow the preferential absorbance of PP (e.g., high ionic strength buffers in the
presence of detergent). The column is eluted under conditions that disrupt antibody/PP binding (e.g.,
a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and

35   PP is collected.

XVI.   Identification of Molecules Which Interact with PP

PP, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PP, washed, and

5 any wells with labeled PP complex are assayed. Data obtained using different concentrations of PP are used to calculate values for the number, affinity, and association of PP with the candidate molecules.

Alternatively, molecules interacting with PP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available

10 kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

15 **XVII.   Demonstration of PP Activity**

PP activity is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). PP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of PP is

20 demonstrated by incubating PP-containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of PP in the assay.

25 In the alternative, PP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM enzyme in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of

30 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × g for 5 min. Acid-soluble [32]Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XVIII. Identification of PP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations

35 along with an appropriate buffer and substrate, as described in the assays in Example XVII. PP

activity is measured for each well and the ability of each compound to inhibit PP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PP activity.

XIX.   Identification of PP Substrates

5   A PP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence. PP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of PP is performed using a method standard in the art or a commercial kit, such as the

10   MUTA-GENE kit from BIO-RAD. For expression of PP mutants in Escherichia coli, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding PP or a glutathione S-transferase (GST)-PP fusion protein. PP mutants are expressed in E. coli and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated

15   transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μg/ml leupeptin/5 μg/ml aprotinin/1 mM benzamidine

20   (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). PP is immunoprecipitated from lysates with an appropriate antibody. GST-PP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

25

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific

30   embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1905692	1	1905692CD1	11	1905692CB1
7476908	2	7476908CD1	12	7476908CB1
7708162	3	7708162CD1	13	7708162CB1
7473603	4	7473603CD1	14	7473603CB1
7476687	5	7476687CD1	15	7476687CB1
7480440	6	7480440CD1	16	7480440CB1
7480570	7	7480570CD1	17	7480570CB1
4555838	8	4555838CD1	18	4555838CB1
636866	9	636866CD1	19	636866CB1
7475576	10	7475576CD1	20	7475576CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	1905692CD1	g3549240	0	[fl][Homo sapiens] putative tyrosine phosphatase (Chen,H. et al. (1999) Hum. Genet. 105 (5), 399-409)
2	7476908CD1	g4325051	1.00E-89	[Mus musculus] putative protein phosphatase type 2C
3	7708162CD1	g3820433	0	[fl][Mus musculus] glucose-6-phosphatase (Arden,S.D. et al. (1999) Diabetes 48 (3), 531-542)
4	7473603CD1	g6692782	2.00E-33	[fl][Homo sapiens] protein phosphatase (Nakamura,K. et al. (1999) Biochem. J. 344 Pt 3, 819-825)
5	7476687CD1	g3123850	2.00E-29	[Homo sapiens] type-2 phosphatidic acid phosphatase alpha-2 (Leung,D.W. et al. (1998) DNA Cell Biol. 17 (4), 377-385)
6	7480440CD1	g12958663	1.00E-142	[fl][Homo sapiens] acid phosphatase variant 3 (Yousef,G.M. et al. (2001) Genomics 74 (3), 385-395)
7	7480570CD1	g13990989	0	[fl][Mus musculus] (AB052157) MAP kinase phosphatase-7
8	4555838CD1	g6650693	1.00E-89	[fl][Homo sapiens] putative tyrosine phosphatase (Uwanogho,D.A. et al. (1999) Genomics 62 (3), 406-416)
9	636866CD1	g3218467	4.00E-93	[fl][Gallus gallus] putative phosphatase (Houston,B. et al. (1999) Biochim. Biophys. Acta 1480, 500-506)
10	7475576CD1	g14031085	1.00E-128	"[fl][Homo sapiens] (AL160175) bA243J16.6 (novel protein with a dual specificity phosphatase, catalytic domain) "

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1905692CD1	545	S224 S254 S373 S383 S446 S495 S64 T231 T243 T336 T355 T368 T371 T453 T481 T536	N450 N485	Transmembrane domain: F94-I114, I129-V147, L161-F178, L193-F209	HMMER
					Tyr_Phosphatase: I330-M342	MOTIFS
2	7476908CD1	360	S187 S201 S254 S266 S309 S354 S355 S51 T134 T167	N117 N252 N343 N353	PROTEIN PHOSPHATASE 2C DM00377|P49597|108-431: V77-L321 DM00377|P49596|1-295: K89-V349 DM00377|P36993|1-304: S97-Q360	BLAST-DOMO
					PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE FAMILY PP2C ISOFORM PD001101: S97-N276	BLAST-PRODOM
					Protein phosphatase 2C p BL01032: R256-D269, P294-D306, C340-V349, Q99-F109, I123-G132, V210-I249	BLIMPS-BLOCKS
					Transmembrane domain: L9-L35,	HMMER
					Protein phosphatase 2C PP2: H91-I344	HMMER-PFAM
					Protein phosphatase 2C Pp2c: I123-G131	MOTIFS
					Signal_cleavage: M1-S37	PROFILESCAN
3	7708162CD1	355	S281 S291 S350 T103 T314	N287 N50 N92	GLUCOSE-6-PHOSPHATASE G6PASE GLYCOGEN BIOSYNTHESIS HYDROLASE TRANSMEMBRANE GLYCOPROTEIN ENDOPLASMIC RETICULUM STORAGE PD014226: M1-Y341	BLAST-PRODOM
					Transmembrane domain: F213-L233	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7473603CD1	493	S136 S280 S316 S356 S362 S393 S410 S417 S470 T172 T231 T258 T351 T475 T91 T97 Y27		VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823|P28562|169-314: E53-M197	BLAST-DOMO
					Transmembrane domain: V152-R174	HMMER
					Dual specificity phosphatase, catalytic domain DSPc: E50-K195	HMMER-PFAM
5	7476687CD1	321	S172 S302 T119 T213	N158	PHOSPHATIDIC ACID PHOSPHATASE HYDROLASE PROTEIN PHOSPHOHYDROLASE TRANSMEMBRANE PHOSPHATIDATE 2A TYPE2 PD005298: V18-G179	BLAST-PRODOM
					Transmembrane domain: Y69-C89, A254-N272	HMMER
					PAP2 superfamily PAP2: T119-G276	HMMER-PFAM
6	7480440CD1	426	S187 S238 S286 S87 T107 Y175	N191 N269 N330 N339	Signal peptide: M1-P28	SPSCAN
					signal peptide: M1-G30	HMMER
					Histidine acid phosphatase: V35-F317	HMMER_PFAM
					Histidine acid phosphatases phosphohistidine proteins BL00616:R40-L47, H288-L301, G297-C312	BLIMPS_BLOCKS
					Histidine acid phosphatases signatures: G8-V75	PROFILESCAN
					Histidine acid phosphatases signatures His_Acid_Phosphat_1: L32-P46 His_Acid_Phosphat_2: M282-A298	MOTIFS
					HYDROLASE ACID PHOSPHATASE 3 PHYTASE MYOINOSITOLHEXAPHOSPHATE 3 PHOSPHOHYDROLASE FAMILY PD003234:V35-L165, L280-P307	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6					SIMILAR TO HISTIDINE ACID PHOSPHATASE FAMILY PD134242:L15-L405	BLAST_PRODOM
					ACID PHOSPHATASE HYDROLASE SIGNAL LYSOSOMAL LAP LYSOSOME PROSTATIC PD002567:P308-C378	BLAST_PRODOM
					MAMMALIAN ACID PHOSPHATASE DM07177|P11117|1-383:M1-I386 DM07177|P15309|1-385:L16-C378	BLAST_DOMO
7	7480570CD1	665	S21 S214 S266 S369 S393 S421 S422 S481 S486 S508 S596 S604 S609 S612 S617 S627 S642 S72 S82 S91 T12 T23 T280 T434 T439 T458 T473 T589	N190 N212 N300 N38 N49 N640	Tyrosine Phosphatase V242-I254	MOTIFS
					Tyrosine specific protein phosphatases active site tyr_phosphatase.prf: L220-I281	PROFILESCAN
					Dual specificity phosphatase, catalytic domain DSPc: G158-K297	HMMER_PFAM
					Rhodanese-like domain Rhodanese: V11-C131	HMMER_PFAM
					Tyrosine specific protein phosphatase BL00383: V242-A252	BLIMPS_BLOCKS
					Dual specificity protein phosphatase PD004588 L15-C170	BLAST_PRODOM
					VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823|A56115|51-336:V85-I298	BLAST_DOMO
8	4555838CD1	254	S124 S245 T142 T28	N209	Transmembrane domain: F210-F229	HMMER
					PROTEIN T15B7.2 SAP185BCK1 INTERGENIC REGION TRANSMEMBRANE PD037603: Y41-L244	BLAST_PRODOM
					Tyrosine specific protein phosphatase signature: L93-V105	MOTIFS
					signal_cleavage: M1-A63	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla-tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	636866CD1	267	S132 S150 S242 S58 T252 T36 T62		PUTATIVE PHOSPHATASE PD123939: M1-K265	BLAST_PRODOM
10	7475576CD1	329	S130 S237 S324 T298	N60	Tyr_Phosphatase V180-I192	MOTIFS
					Tyrosine specific protein phosphatases active site tyr_phosphatase.prf: K159-A219	PROFILESCAN
					Dual specificity phosphatase, catalytic domain DSPc: G98-A235	HMMER_PFAM
					Tyrosine specific protein phosphatase BL00383:V180-T190	BLIMPS_BLOCKS
					VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823 I38890 29-320: P92-F232 DM03823 A56115 51-336: A78-F232 DM03823 Q02256 1-174: G176-F232	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
11	1905692CB1	1845	686-1123	g2114900	1372	1845
				6024861H1 (TESTNOT11)	1	305
				g6450924	1359	1834
				FL1905692_g7717240_g3549240	47	1675
12	7476908CB1	2451	1797-1864, 1159-1372, 1952-2451	7190768 3V1	1348	2023
				55061893J1	1	613
				71906713V1	1661	2451
				6869964H1 (BRAGNON02)	607	1321
				71908193V1	1284	1995
				4874491H1 (COLDNOT01)	563	821
13	7708162CB1	1105	1-32, 258-1105	55098203H1	1	110
				FL7708162_g8469075_0000 41_g3820433	38	1105
14	7473603CB1	1730	1680-1730, 599-1535	FL7473603_g7329576_0000 27_g6692780	1	897
				8124012H1 (HEAONOC01)	1033	1730
				8123914H1 (HEAONOC01)	649	1405
15	7476687CB1	2145	65-115, 1411-1810	70317889D1	963	1544
				55093694H1	766	1033
				2705182H1 (PONSAZT01)	540	808
				55077827H1	242	700
				g876492	1811	2145
				2705182T6 (PONSAZT01)	1242	1860
				6857089H1 (BRAIFEN08)	1658	1869
				g6301939	1	416
16	7480440CB1	1352	333-551, 1-74, 963-982	g2953599	1060	1352
				g2953528	995	1342
				GNN.g8575999_000022_004	1	1281

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
17	7480570CB1	3766	2430-2650, 1-709, 1531-1584, 1298-1393	7321002H1 (SPLNFEE03)	1332	2025
				1234795T6 (LUNGFET03)	2115	2622
				3232147F6 (COLNUCT03)	1	553
				210522R6 (SPLNNOT02)	3001	3473
				760852SJ1 (COLRTUE01)	1584	2180
				55116305J1	2170	2994
				7389826H2 (LIVRFEE02)	955	1407
				2235159F6 (PANCTUT02)	3316	3766
				4155374T6 (ADRENOT14)	2775	3470
				55098593J1	474	1332
18	4555838CB1	2656	1-138, 2166-2656	7620770J1 (HEARFEE03)	1157	1726
				6883412H1 (BRAHTDR03)	2026	2656
				70851175V1	1798	2510
				7273895H1 (PROSUNJ01)	672	1207
				7326636H2 (SPLNTUE01)	1	595
				6911394J1 (PITUDIR01)	1317	1909
				6850290H1 (BRAIFEN08)	372	1033
19	636866CB1	1292	1-131, 654-690	g6656936	691	1135
				g3412425	952	1286
				636866H1 (NEUTGMT01)	840	1117
				2433115H1 (BRAVUNT02)	1	270
				g3330652	867	1285
				7675126J2 (NOSETUE01)	298	847
				6770613H1 (BRAUNOR01)	655	1048
				656517H1 (EOSINOT03)	123	402
20	7475576CB1	1325	1-185, 1217-1277	71702036V1	790	1325
				55032319H1	186	930
				GNN.g9187276_000003_008.edit	1	545

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
11	1905692CB1	TESTNOT11
12	7476908CB1	BRALNOT01
13	7708162CB1	PANCNOE02
14	7473603CB1	HEAONOC01
15	7476687CB1	PONSAZT01
17	7480570CB1	LUNGFET03
18	4555838CB1	COLNPOT01
19	636866CB1	NOSETUE01
20	7475576CB1	TESTNOT17

Table 6

Library	Vector	Library Description
BRALNOT01	pINCY	Library was constructed using RNA isolated from thalamus tissue removed from a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
COLNPOT01	pINCY	Library was constructed using RNA isolated from colon polyp tissue removed from a 40-year-old Caucasian female during a total colectomy. Pathology indicated an inflammatory pseudopolyp; this tissue was associated with a focally invasive grade 2 adenocarcinoma and multiple tubulivillous adenomas. Patient history included a benign neoplasm of the bowel.
HEAONOC01	PSPORT1	This large size fractionated library was constructed using RNA isolated from the aorta of a 39-year-old Caucasian male, who died from a gunshot wound. Serology was positive for cytomegalovirus (CMV). Patient history included tobacco abuse (one pack of cigarettes per day for 25 years), and occasionally cocaine, marijuana, and alcohol use.
LUNGFET03	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
NOSETUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from nasal and cribriform tumor tissue removed from a 45-year-old Caucasian male during total face ostectomy with reconstruction, rhinotomy and craniotomy. Pathology indicated olfactory neuroblastoma in the nasal cavity and cribriform region. The patient presented with cancer of the head, face and neck, and epistaxis. Patient history included extrinsic asthma, cancer of the head, face and neck, and epistaxis. Previous surgeries included total face ostectomy with reconstruction. Patient medications included Biaxin, Atessalon, and Valium. The patient received radiation treatments. Family history included chronic sinusitis in the mother and type II diabetes in the father.
PANCNOE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from pancreatic tissue removed from an 8-year-old Black male, who died from anoxia. Serologies were negative. Patient medications included DDAVP, Versed, and labetalol.
PONSAZT01	pINCY	Library was constructed using RNA isolated from diseased pons tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
TESTNOT11	pINCY	Library was constructed using RNA isolated from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. Patient history included drug use (tobacco, marijuana, and cocaine use), and medications included Lithium, Ritalin, and Paxil.

Table 6 (cont.)

Library	Vector	Library Description
TESTNOT17	pINCY	Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs*: Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs*: fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences*: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits*: Probability value= 1.0E-3 or less *Signal peptide hits*: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al, eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of

5   SEQ ID NO:1-10,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical

to an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected

from the group consisting of SEQ ID NO:1-10, and

10   d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from

the group consisting of SEQ ID NO:1-10.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-

10.

15

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

20   5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID

NO:11-20.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a

polynucleotide of claim 3.

25

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30   9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said

cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide

comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of

claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

5   11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20,

10   c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a

15   polynucleotide of claim 11.

13. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

20   comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

25

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

30   a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected
5   from the group consisting of SEQ ID NO:1-10.

18. A method for treating a disease or condition associated with decreased expression of functional PP, comprising administering to a patient in need of such treatment the composition of claim 16.
10

19. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
    a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
    b) detecting agonist activity in the sample.
15

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of
20   functional PP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
25   a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
    b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.
30

24. A method for treating a disease or condition associated with overexpression of functional PP, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, the method comprising:

    a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

    b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

5   compound that specifically binds to the polypeptide of claim 1.

    26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

    a) combining the polypeptide of claim 1 with at least one test compound under conditions

10   permissive for the activity of the polypeptide of claim 1,

    b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

    c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change

15   in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

    27. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

20   comprising:

    a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

    b) detecting altered expression of the target polynucleotide, and

    c) comparing the expression of the target polynucleotide in the presence of varying amounts

25   of the compound and in the absence of the compound.

    28. A method of assessing toxicity of a test compound, the method comprising:

    a) treating a biological sample containing nucleic acids with the test compound,

    b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at

30   least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof,

    c) quantifying the amount of hybridization complex, and

35   d) comparing the amount of hybridization complex in the treated biological sample with the

amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5   29. A diagnostic test for a condition or disease associated with the expression of PP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence

10   of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

15   c) a Fab fragment,

d) a F(ab')$_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

20

32. A method of diagnosing a condition or disease associated with the expression of PP in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

25

34. A method of diagnosing a condition or disease associated with the expression of PP in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

30   10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

95

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in the sample.

5   44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 from a sample, the method comprising:

      a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

      b) separating the antibody from the sample and obtaining the purified polypeptide having an 10 amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

      45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

      46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

15

      47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

      48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20   49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

      50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

      51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

25

      52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

      53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30   54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

      55. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:11.

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:12.

57. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:13.

58. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:14.

59. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:15.

60. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:16.

61. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:17.

62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:18.

63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:19.

64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.

<110> INCYTE GENOMICS, INC.
      TANG, Y.Tom
      ELLIOTT, Vicki S.
      RAMKUMAR, Jayalaxmi
      YAO, Monique G.
      BURFORD, Neil
      WANG, Yumei E.
      STEWART, Elizabeth A.
      GANDHI, Ameena R.
      PATTERSON, Chandra
      LEE, Ernestine A.
      HAFALIA, April J.A.
      LU, Dyung Aina M.
      TRIBOULEY, Catherine M.
      GRIFFIN, Jennifer A.
      BAUGHN, Mariah R.
      YUE, Henry
      WARREN, Bridget A.
      NGUYEN, Danniel B.
      WALIA, Narinder K.
      KEARNEY, Liam

<120> PROTEIN PHOSPHATASES

<130> PI-0173 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/221,679; 60/223,272; 60/224,309; 60/226,728; 60/229,254;
      60/231,366
<160> 2000-07-28; 2000-08-03; 2000-08-10; 2000-08-18; 2000-08-30;
      2000-09-08

<170> PERL Program

<210> 1
<211> 545
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1905692CD1

<400> 1
Met Asn Glu Ser Pro Asp Pro Thr Asp Leu Ala Gly Val Ile Ile
 1   5   10   15
Glu Leu Gly Pro Asn Asp Ser Pro Gln Thr Ser Glu Phe Lys Gly
                20   25   30
Ala Thr Glu Glu Ala Pro Ala Lys Glu Ser Pro His Thr Ser Glu
                35   40   45
Phe Lys Gly Ala Ala Arg Val Ser Pro Ile Ser Glu Ser Val Leu
                50   55   60
Ala Arg Leu Ser Lys Phe Glu Asp Glu Asp Ala Glu Asn Val Ala
                65   70   75
Ser Tyr Asp Ser Lys Ile Lys Lys Ile Val His Ser Ile Val Ser
                80   85   90
Ser Phe Ala Phe Gly Leu Phe Gly Val Phe Leu Val Leu Leu Asp
                95   100   105
Val Thr Leu Ile Leu Ala Asp Leu Ile Phe Thr Asp Ser Lys Leu
                110   115   120
Tyr Ile Pro Leu Glu Tyr Arg Ser Ile Ser Leu Ala Ile Ala Leu

```
                   125                  130                  135
Phe Phe Leu Met Asp Val Leu Leu Arg Val Phe Val Glu Arg Arg
                   140                  145                  150
Gln Gln Tyr Phe Ser Asp Leu Phe Asn Ile Leu Asp Thr Ala Ile
                   155                  160                  165
Ile Val Ile Leu Leu Leu Val Asp Val Val Tyr Ile Phe Phe Asp
                   170                  175                  180
Ile Lys Leu Leu Arg Asn Ile Pro Arg Trp Thr His Leu Leu Arg
                   185                  190                  195
Leu Leu Arg Leu Ile Ile Leu Leu Arg Ile Phe His Leu Phe His
                   200                  205                  210
Gln Lys Arg Gln Leu Glu Lys Leu Ile Arg Arg Arg Val Ser Glu
                   215                  220                  225
Asn Lys Arg Arg Tyr Thr Arg Asp Gly Phe Asp Leu Asp Leu Thr
                   230                  235                  240
Tyr Val Thr Glu Arg Ile Ile Ala Met Ser Phe Pro Ser Ser Gly
                   245                  250                  255
Arg Gln Ser Phe Tyr Arg Asn Pro Ile Lys Val Ile Pro Tyr Arg
                   260                  265                  270
Asp Met Thr Tyr Ile Leu Phe Ile Leu Gly Glu Arg Ala Tyr Asp
                   275                  280                  285
Pro Lys His Phe His Asn Arg Val Val Arg Ile Met Ile Asp Asp
                   290                  295                  300
His Asn Val Pro Thr Leu His Gln Met Val Val Phe Thr Lys Glu
                   305                  310                  315
Val Asn Glu Trp Met Ala Gln Asp Leu Glu Asn Ile Val Ala Ile
                   320                  325                  330
His Cys Lys Gly Gly Thr Asp Arg Thr Gly Thr Met Val Cys Ala
                   335                  340                  345
Phe Leu Ile Ala Ser Glu Ile Cys Ser Thr Ala Lys Glu Ser Leu
                   350                  355                  360
Tyr Tyr Phe Gly Glu Arg Arg Thr Asp Lys Thr His Ser Glu Lys
                   365                  370                  375
Phe Gln Gly Val Lys Thr Pro Ser Gln Lys Arg Tyr Val Ala Tyr
                   380                  385                  390
Phe Ala Gln Val Lys His Leu Tyr Asn Trp Asn Leu Pro Pro Arg
                   395                  400                  405
Arg Ile Leu Phe Ile Lys His Phe Ile Ile Tyr Ser Ile Pro Arg
                   410                  415                  420
Tyr Val Arg Asp Leu Lys Ile Gln Ile Glu Met Glu Lys Lys Val
                   425                  430                  435
Val Phe Ser Thr Ile Ser Leu Gly Lys Cys Ser Val Leu Asp Asn
                   440                  445                  450
Ile Thr Thr Asp Lys Ile Leu Ile Asp Val Phe Asp Gly Pro Pro
                   455                  460                  465
Leu Tyr Asp Asp Val Lys Val Gln Phe Phe Ser Ser Asn Leu Pro
                   470                  475                  480
Thr Tyr Tyr Asp Asn Cys Ser Phe Tyr Phe Trp Leu His Thr Ser
                   485                  490                  495
Phe Ile Glu Asn Asn Arg Leu Tyr Leu Pro Lys Asn Glu Leu Asp
                   500                  505                  510
Asn Leu His Lys Gln Lys Ala Arg Arg Ile Tyr Pro Ser Asp Phe
                   515                  520                  525
Ala Val Glu Ile Leu Phe Gly Glu Lys Met Thr Ser Ser Asp Val
                   530                  535                  540
Val Ala Gly Ser Asp
                   545
```

<210> 2
<211> 360
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476908CD1

<400> 2
Met Ile Glu Asp Thr Met Thr Leu Leu Ser Leu Leu Gly Arg Ile
 1   5   10   15
Met Arg Tyr Phe Leu Leu Arg Pro Glu Thr Leu Phe Leu Leu Cys
                20   25   30
Ile Ser Leu Ala Leu Trp Ser Tyr Phe Phe His Thr Asp Glu Val
                35   40   45
Lys Thr Ile Val Lys Ser Ser Arg Asp Ala Val Lys Met Val Lys
                50   55   60
Ser Lys Val Ala Glu Thr Met Gln Asn Asp Arg Leu Gly Gly Leu
                65   70   75
Asp Val Leu Glu Ala Glu Phe Ser Lys Thr Trp Glu Phe Lys Asn
                80   85   90
His Asn Val Ala Val Tyr Ser Ile Gln Gly Arg Arg Asp His Met
                95   100   105
Glu Asp Arg Phe Glu Val Leu Thr Asp Leu Ala Asn Lys Thr His
                110   115   120
Pro Ser Ile Phe Gly Ile Phe Asp Gly His Gly Gly Glu Thr Ala
                125   130   135
Ala Glu Tyr Val Lys Ser Arg Leu Pro Glu Ala Leu Lys Gln His
                140   145   150
Leu Gln Asp Tyr Glu Lys Asp Lys Glu Asn Ser Val Leu Ser Tyr
                155   160   165
Gln Thr Ile Leu Glu Gln Gln Ile Leu Ser Ile Asp Arg Glu Met
                170   175   180
Leu Glu Lys Leu Thr Val Ser Tyr Asp Glu Ala Gly Thr Thr Cys
                185   190   195
Leu Ile Ala Leu Leu Ser Asp Lys Asp Leu Thr Val Ala Asn Val
                200   205   210
Gly Asp Ser Arg Gly Val Leu Cys Asp Lys Asp Gly Asn Ala Ile
                215   220   225
Pro Leu Ser His Asp His Lys Pro Tyr Gln Leu Lys Glu Arg Lys
                230   235   240
Arg Ile Lys Arg Ala Gly Gly Phe Ile Ser Phe Asn Gly Ser Trp
                245   250   255
Arg Val Gln Gly Ile Leu Ala Met Ser Arg Ser Leu Gly Asp Tyr
                260   265   270
Pro Leu Lys Asn Leu Asn Val Val Ile Pro Asp Pro Asp Ile Leu
                275   280   285
Thr Phe Asp Leu Asp Lys Leu Gln Pro Glu Phe Met Ile Leu Ala
                290   295   300
Ser Asp Gly Leu Trp Asp Ala Phe Ser Asn Glu Glu Ala Val Arg
                305   310   315
Phe Ile Lys Glu Arg Leu Asp Glu Pro His Phe Gly Ala Lys Ser
                320   325   330
Ile Val Leu Gln Ser Phe Tyr Arg Gly Cys Pro Asp Asn Ile Thr
                335   340   345
Val Met Val Val Lys Phe Arg Asn Ser Ser Lys Thr Glu Glu Gln
                350   355   360

<210> 3
<211> 355
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7708162CD1

```
<400> 3
Met Asp Phe Leu His Arg Asn Gly Val Leu Ile Ile Gln His Leu
1               5                   10                  15
Gln Lys Asp Tyr Arg Ala Tyr Tyr Thr Phe Leu Asn Phe Met Ser
            20                  25                  30
Asn Val Gly Asp Pro Arg Asn Ile Phe Phe Ile Tyr Phe Pro Leu
        35                  40                  45
Cys Phe Gln Phe Asn Gln Thr Val Gly Thr Lys Met Ile Trp Val
            50                  55                  60
Ala Val Ile Gly Asp Trp Leu Asn Leu Ile Phe Lys Trp Ile Leu
65                  70                  75
Phe Gly His Arg Pro Tyr Trp Trp Val Gln Glu Thr Gln Ile Tyr
            80                  85                  90
Pro Asn His Ser Ser Pro Cys Leu Glu Gln Phe Pro Thr Thr Cys
            95                  100                 105
Glu Thr Gly Pro Gly Ser Pro Ser Gly His Ala Met Gly Ala Ser
            110                 115                 120
Cys Val Trp Tyr Val Met Val Thr Ala Ala Leu Ser His Thr Val
            125                 130                 135
Cys Gly Met Asp Lys Phe Ser Ile Thr Leu His Arg Leu Thr Trp
            140                 145                 150
Ser Phe Leu Trp Ser Val Phe Trp Leu Ile Gln Ile Ser Val Cys
            155                 160                 165
Ile Ser Arg Val Phe Ile Ala Thr His Phe Pro His Gln Val Ile
            170                 175                 180
Leu Gly Val Ile Gly Gly Met Leu Val Ala Glu Ala Phe Glu His
            185                 190                 195
Thr Pro Gly Ile Gln Thr Ala Ser Leu Gly Thr Tyr Leu Lys Thr
            200                 205                 210
Asn Leu Phe Leu Phe Leu Phe Ala Val Gly Phe Tyr Leu Leu Leu
            215                 220                 225
Arg Val Leu Asn Ile Asp Leu Leu Trp Ser Val Pro Ile Ala Lys
            230                 235                 240
Lys Trp Cys Ala Asn Pro Asp Trp Ile His Ile Asp Thr Thr Pro
            245                 250                 255
Phe Ala Gly Leu Val Arg Asn Leu Gly Val Leu Phe Gly Leu Gly
            260                 265                 270
Phe Ala Ile Asn Ser Glu Met Phe Leu Leu Ser Cys Arg Gly Gly
            275                 280                 285
Asn Asn Tyr Thr Leu Ser Phe Arg Leu Leu Cys Ala Leu Thr Ser
            290                 295                 300
Leu Thr Ile Leu Gln Leu Tyr His Phe Leu Gln Ile Pro Thr His
            305                 310                 315
Glu Glu His Leu Phe Tyr Val Leu Ser Phe Cys Lys Ser Ala Ser
            320                 325                 330
Ile Pro Leu Thr Val Val Ala Phe Ile Pro Tyr Ser Val His Met
            335                 340                 345
Leu Met Lys Gln Ser Gly Lys Lys Ser Gln
            350                 355

<210> 4
<211> 493
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473603CD1

<400> 4
Met Leu Glu Ser Ala Glu Gln Leu Leu Val Glu Asp Leu Tyr Asn
1               5                   10                  15
Arg Val Arg Glu Lys Met Asp Asp Thr Ser Leu Tyr Asn Thr Pro
```

```
                      20                      25                      30
Cys Val Leu Asp Leu Gln Arg Ala Leu Val Gln Asp Arg Gln Glu
                 35                      40           .           45
Ala Pro Trp Asn Glu Val Asp Glu Val Trp Pro Asn Val Phe Ile
                 50                      55                      60
Ala Asp Arg Ser Val Ala Val Asn Lys Gly Arg Leu Lys Arg Leu
                 65                      70                      75
Gly Ile Thr His Ile Leu Asn Ala Ala His Gly Thr Gly Val Tyr
                 80                      85                      90
Thr Gly Pro Glu Phe Tyr Thr Gly Leu Glu Ile Gln Tyr Leu Gly
                 95                     100                     105
Val Glu Val Asp Asp Phe Pro Glu Val Asp Ile Ser Gln His Phe
                110                     115                     120
Arg Lys Ala Tyr Cys His Tyr Ile Ile Phe Ser Cys Val Phe Ile
                125                     130                     135
Ser Gly Lys Val Leu Val Ser Ser Glu Met Gly Ile Ser Arg Ser
    .           140                     145                     150
Ala Val Leu Val Val Ala Tyr Leu Met Ile Phe His Asn Met Ala
                155                     160 .                   165
Ile Leu Glu Ala Leu Met Thr Val Arg Lys Lys Arg Ala Ile Tyr
                170                     175                     180
Pro Asn Glu Gly Phe Leu Lys Gln Leu Arg Glu Leu Asn Glu Lys
                185                     190                     195
Leu Met Glu Glu Arg Glu Glu Asp Tyr Gly Arg Glu Gly Gly Ser
                200                     205                     210
Ala Glu Ala Glu Glu Gly Glu Gly Thr Gly Ser Met Leu Gly Ala
                215                     220                     225
Arg Val His Ala Leu Thr Val Glu Glu Glu Asp Asp Ser Ala Ser
                230                     235                     240
His Leu Ser Gly Ser Ser Leu Gly Lys Ala Thr Gln Ala Ser Lys
                245                     250                     255
Pro Leu Thr Leu Ile Asp Glu Glu Glu Glu Lys Leu Tyr Glu
                260                     265                     270
Gln Trp Lys Lys Gly Gln Gly Leu Leu Ser Asp Lys Val Pro Gln
                275                     280                     285
Asp Gly Gly Gly Trp Arg Ser Ala Ser Ser Gly Gln Gly Gly Glu
                290                     295                     300
Glu Leu Glu Asp Glu Asp Val Glu Arg Ile Ile Gln Glu Trp Gln
                305                     310                     315
Ser Arg Asn Glu Arg Tyr Gln Ala Glu Gly Tyr Arg Arg Trp Gly
                320                     325 .                   330
Arg Glu Glu Glu Lys Glu Glu Glu Ser Asp Ala Gly Ser Ser Val
                335                     340                     345
Gly Arg Arg Arg Arg Thr Leu Ser Glu Ser Ser Ala Trp Glu Ser
                350                     355                     360
Val Ser Ser His Asp Ile Trp Val Leu Lys Gln Gln Leu Glu Leu
                365                     370                     375
Asn Arg Pro Asp His Gly Arg Arg Arg Ala Asp Ser Met Ser
                380                     385                     390
Ser Glu Ser Thr Trp Asp Ala Trp Asn Glu Arg Leu Leu Glu Ile
    .           395                     400                     405
Glu Lys Glu Ala Ser Arg Arg Tyr His Ala Lys Ser Lys Arg Glu
                410                     415                     420
Glu Ala Ala Asp Arg Ser Ser Glu Ala Gly Ser Arg Val Arg Glu
                425                     430                     435
Asp Asp Glu Asp Ser Val Gly Ser Glu Ala Ser Ser Phe Tyr Asn
                440                     445                     450
Phe Cys Ser Arg Asn Lys Asp Lys Leu Thr Ala Trp Lys Asp Gly
                455             . .     460                     465
Arg Ser Arg Glu Ser Asn Leu Asp Phe Thr Arg Lys Thr Trp Glu
                470                     475                     480
Arg Glu Thr Ala Ala Val Ser Pro Val Gln Arg Arg Gln
                485                     490
```

```
<210> 5
<211> 321
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476687CD1

<400> 5
Met Pro Leu Leu Pro Ala Ala Leu Thr Ser Ser Met Leu Tyr Phe
 1               5                  10                  15
Gln Met Val Ile Met Ala Gly Thr Val Met Leu Ala Tyr Tyr Phe
                20                  25                  30
Glu Tyr Thr Asp Thr Phe Thr Val Asn Val Gln Gly Phe Phe Cys
                35                  40                  45
His Asp Ser Ala Tyr Arg Lys Pro Tyr Pro Gly Pro Glu Asp Ser
                50                  55                  60
Ser Ala Val Pro Pro Val Leu Leu Tyr Ser Leu Ala Ala Gly Val
                65                  70                  75
Pro Val Leu Val Ile Ile Val Gly Glu Thr Ala Val Phe Cys Leu
                80                  85                  90
Gln Leu Ala Thr Arg Asp Phe Glu Asn Gln Glu Lys Thr Ile Leu
                95                  100                 105
Thr Gly Asp Cys Cys Tyr Ile Asn Pro Leu Val Arg Arg Thr Val
                110                 115                 120
Arg Phe Leu Gly Ile Tyr Thr Phe Gly Leu Phe Ala Thr Asp Ile
                125                 130                 135
Phe Val Asn Ala Gly Gln Val Val Thr Gly Asn Leu Ala Pro His
                140                 145                 150
Phe Leu Ala Leu Cys Lys Pro Asn Tyr Thr Ala Leu Gly Cys Gln
                155                 160                 165
Gln Tyr Thr Gln Phe Ile Ser Gly Glu Glu Ala Cys Thr Gly Asn
                170                 175                 180
Pro Asp Leu Ile Met Arg Ala Arg Lys Thr Phe Pro Ser Lys Glu
                185                 190                 195
Ala Ala Leu Ser Val Tyr Ala Ala Met Tyr Leu Thr Met Tyr Ile
                200                 205                 210
Thr Asn Thr Ile Lys Ala Lys Gly Thr Arg Leu Ala Lys Pro Val
                215                 220                 225
Leu Cys Leu Gly Leu Met Cys Leu Ala Phe Leu Thr Gly Leu Asn
                230                 235                 240
Arg Val Ala Glu Tyr Arg Asn His Trp Ser Asp Val Ile Ala Gly
                245                 250                 255
Phe Leu Val Gly Ile Ser Ile Ala Val Phe Leu Val Val Cys Val
                260                 265                 270
Val Asn Asn Phe Lys Gly Arg Gln Ala Glu Asn Glu His Ile His
                275                 280                 285
Met Asp Asn Leu Ala Gln Met Pro Met Ile Ser Ile Pro Arg Val
                290                 295                 300
Glu Ser Pro Leu Glu Lys Val Thr Ser Val Gln Asn His Ile Thr
                305                 310                 315
Ala Phe Ala Glu Val Thr
                320

<210> 6
<211> 426
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480440CD1
```

<400> 6

```
Met Ala Gly Leu Gly Phe Trp Gly His Pro Ala Gly Pro Leu Leu
 1               5                  10                  15
Leu Leu Leu Leu Leu Val Leu Pro Pro Arg Ala Leu Pro Glu Gly
                20                  25                  30
Pro Leu Val Phe Val Ala Leu Val Phe Arg His Gly Asp Arg Ala
                35                  40                  45
Pro Leu Ala Ser Tyr Pro Met Asp Pro His Lys Glu Val Ala Ser
                50                  55                  60
Thr Leu Trp Pro Arg Gly Leu Gly Gln Leu Thr Thr Glu Gly Val
                65                  70                  75
Arg Gln Gln Leu Glu Leu Gly Arg Phe Leu Arg Ser Arg Tyr Glu
                80                  85                  90
Ala Phe Leu Ser Pro Glu Tyr Arg Arg Glu Glu Val Tyr Ile Arg
                95                 100                 105
Ser Thr Asp Phe Asp Arg Thr Leu Glu Ser Ala Gln Ala Asn Leu
               110                 115                 120
Ala Gly Leu Phe Pro Glu Ala Ala Pro Gly Ser Pro Glu Ala Arg
               125                 130                 135
Trp Arg Pro Ile Pro Val His Thr Val Pro Val Ala Glu Asp Lys
               140                 145                 150
Leu Leu Arg Phe Pro Met Arg Ser Cys Pro Arg Tyr His Glu Leu
               155                 160                 165
Leu Arg Glu Ala Thr Glu Ala Ala Glu Tyr Gln Glu Ala Leu Glu
               170                 175                 180
Gly Trp Thr Gly Phe Leu Ser Arg Leu Glu Asn Phe Thr Gly Leu
               185                 190                 195
Ser Leu Val Gly Glu Pro Leu Arg Arg Ala Trp Lys Val Leu Asp
               200                 205                 210
Thr Leu Met Cys Gln Gln Ala His Gly Leu Pro Leu Pro Ala Trp
               215                 220                 225
Ala Ser Pro Asp Val Leu Arg Thr Leu Ala Gln Ile Ser Ala Leu
               230                 235                 240
Asp Ile Gly Ala His Val Gly Pro Pro Arg Ala Ala Glu Lys Ala
               245                 250                 255
Gln Leu Thr Gly Gly Ile Leu Leu Asn Ala Ile Leu Ala Asn Phe
               260                 265                 270
Ser Arg Val Gln Arg Leu Gly Leu Pro Leu Lys Met Val Met Tyr
               275                 280                 285
Ser Ala His Asp Ser Thr Leu Leu Ala Leu Gln Gly Ala Leu Gly
               290                 295                 300
Leu Tyr Asp Gly His Thr Pro Pro Tyr Ala Ala Cys Leu Gly Phe
               305                 310                 315
Glu Phe Arg Lys His Leu Gly Asn Pro Ala Lys Asp Gly Gly Asn
               320                 325                 330
Val Thr Val Ser Leu Phe Tyr Arg Asn Asp Ser Ala His Leu Pro
               335                 340                 345
Leu Pro Leu Ser Leu Pro Gly Cys Pro Ala Pro Cys Pro Leu Gly
               350                 355                 360
Arg Phe Tyr Gln Leu Thr Ala Pro Ala Arg Pro Pro Ala His Gly
               365                 370                 375
Val Ser Cys His Gly Pro Tyr Glu Ala Ala Ile Pro Pro Ala Pro
               380                 385                 390
Val Val Pro Leu Leu Ala Gly Ala Val Ala Val Leu Val Ala Leu
               395                 400                 405
Ser Leu Gly Leu Gly Leu Leu Ala Trp Arg Pro Gly Cys Leu Arg
               410                 415                 420
Ala Leu Gly Gly Pro Val
               425
```

<210> 7
<211> 665
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480570CD1

<400> 7

```
Met Ala His Glu Met Ile Gly Thr Gln Ile Val Thr Glu Arg Leu
 1               5                  10                  15
Val Ala Leu Leu Glu Ser Gly Thr Glu Lys Val Leu Leu Ile Asp
                20                  25                  30
Ser Arg Pro Phe Val Glu Tyr Asn Thr Ser His Ile Leu Glu Ala
                35                  40                  45
Ile Asn Ile Asn Cys Ser Lys Leu Met Lys Arg Arg Leu Gln Gln
                50                  55                  60
Asp Lys Val Leu Ile Thr Glu Leu Ile Gln His Ser Ala Lys His
                65                  70                  75
Lys Val Asp Ile Asp Cys Ser Gln Lys Val Val Val Tyr Asp Gln
                80                  85                  90
Ser Ser Gln Asp Val Ala Ser Leu Ser Ser Asp Cys Phe Leu Thr
                95                 100                 105
Val Leu Leu Gly Lys Leu Glu Lys Ser Phe Asn Ser Val His Leu
               110                 115                 120
Leu Ala Gly Gly Phe Ala Glu Phe Ser Arg Cys Phe Pro Gly Leu
               125                 130                 135
Cys Glu Gly Lys Ser Thr Leu Val Pro Thr Cys Ile Ser Gln Pro
               140                 145                 150
Cys Leu Pro Val Ala Asn Ile Gly Pro Thr Arg Ile Leu Pro Asn
               155                 160                 165
Leu Tyr Leu Gly Cys Gln Arg Asp Val Leu Asn Lys Glu Leu Met
               170                 175                 180
Gln Gln Asn Gly Ile Gly Tyr Val Leu Asn Ala Ser Asn Thr Cys
               185                 190                 195
Pro Lys Pro Asp Phe Ile Pro Glu Ser His Phe Leu Arg Val Pro
               200                 205                 210
Val Asn Asp Ser Phe Cys Glu Lys Ile Leu Pro Trp Leu Asp Lys
               215                 220                 225
Ser Val Asp Phe Ile Glu Lys Ala Lys Ala Ser Asn Gly Cys Val
               230                 235                 240
Leu Val His Cys Leu Ala Gly Ile Ser Arg Ser Ala Thr Ile Ala
               245                 250                 255
Ile Ala Tyr Ile Met Lys Arg Met Asp Met Ser Leu Asp Glu Ala
               260                 265                 270
Tyr Arg Phe Val Lys Glu Lys Arg Pro Thr Ile Ser Pro Asn Phe
               275                 280                 285
Asn Phe Leu Gly Gln Leu Leu Asp Tyr Glu Lys Lys Ile Lys Asn
               290                 295                 300
Gln Thr Gly Ala Ser Gly Pro Lys Ser Lys Leu Lys Leu Leu His
               305                 310                 315
Leu Glu Lys Pro Asn Glu Pro Val Pro Ala Val Ser Glu Gly Gly
               320                 325                 330
Gln Lys Ser Glu Thr Pro Leu Ser Pro Pro Cys Ala Asp Ser Ala
               335                 340                 345
Thr Ser Glu Ala Ala Gly Gln Arg Pro Val His Pro Ala Ser Val
               350                 355                 360
Pro Ser Val Pro Ser Val Gln Pro Ser Leu Leu Glu Asp Ser Pro
               365                 370                 375
Leu Val Gln Ala Leu Ser Gly Leu His Leu Ser Ala Asp Arg Leu
               380                 385                 390
Glu Asp Ser Asn Lys Leu Lys Arg Ser Phe Ser Leu Asp Ile Lys
               395                 400                 405
Ser Val Ser Tyr Ser Ala Ser Met Ala Ala Ser Leu His Gly Phe
               410                 415                 420
```

```
Ser Ser Ser Glu Asp Ala Leu Glu Tyr Tyr Lys Pro Ser Thr Thr
            425                 430                 435
Leu Asp Gly Thr Asn Lys Leu Cys Gln Phe Ser Pro Val Gln Glu
            440                 445                 450
Leu Ser Glu Gln Thr Pro Glu Thr Ser Pro Asp Lys Glu Glu Ala
            455                 460                 465
Ser Ile Pro Lys Lys Leu Gln Thr Ala Arg Pro Ser Asp Ser Gln
            470                 475                 480
Ser Lys Arg Leu His Ser Val Arg Thr Ser Ser Ser Gly Thr Ala
            485                 490                 495
Gln Arg Ser Leu Leu Ser Pro Leu His Arg Ser Gly Ser Val Glu
            500                 505                 510
Asp Asn Tyr His Thr Ser Phe Leu Phe Gly Leu Ser Thr Ser Gln
            515                 520                 525
Gln His Leu Thr Lys Ser Ala Gly Leu Gly Leu Lys Gly Trp His
            530                 535                 540
Ser Asp Ile Leu Ala Pro Gln Thr Ser Thr Pro Ser Leu Thr Ser
            545                 550                 555
Ser Trp Tyr Phe Ala Thr Glu Ser Ser His Phe Tyr Ser Ala Ser
            560                 565                 570
Ala Ile Tyr Gly Gly Ser Ala Ser Tyr Ser Ala Tyr Ser Cys Ser
            575                 580                 585
Gln Leu Pro Thr Cys Gly Asp Gln Val Tyr Ser Val Arg Arg Arg
            590                 595                 600
Gln Lys Pro Ser Asp Arg Ala Asp Ser Arg Arg Ser Trp His Glu
            605                 610                 615
Glu Ser Pro Phe Glu Lys Gln Phe Lys Arg Arg Ser Cys Gln Met
            620                 625                 630
Glu Phe Gly Glu Ser Ile Met Ser Glu Asn Arg Ser Arg Glu Glu
            635                 640                 645
Leu Gly Lys Val Gly Ser Gln Ser Ser Phe Ser Gly Ser Met Glu
            650                 655                 660
Ile Ile Glu Val Ser
            665
```

```
<210> 8
<211> 254
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4555838CD1

<400> 8
Met Ala Ala Val Ala Ala Thr Ala Ala Ala Lys Gly Asn Gly Gly
  1               5                  10                  15
Gly Gly Gly Arg Ala Gly Ala Gly Asp Ala Ser Gly Thr Arg Lys
            20                  25                  30
Lys Lys Gly Pro Gly Pro Pro Ala Thr Ala Tyr Leu Val Ile Tyr
            35                  40                  45
Asn Val Val Met Thr Ala Gly Trp Leu Val Ile Ala Val Gly Leu
            50                  55                  60
Val Arg Ala Tyr Leu Ala Lys Gly Ser Tyr His Ser Leu Tyr Tyr
            65                  70                  75
Ser Ile Glu Lys Pro Leu Lys Phe Phe Gln Thr Gly Ala Leu Leu
            80                  85                  90
Glu Ile Leu His Cys Ala Ile Gly Ile Val Pro Ser Ser Val Val
            95                  100                 105
Leu Thr Ser Phe Gln Val Met Ser Arg Val Phe Leu Ile Trp Ala
            110                 115                 120
Val Thr His Ser Val Lys Glu Val Gln Ser Glu Asp Ser Val Leu
            125                 130                 135
```

```
Leu Phe Val Ile Ala Trp Thr Ile Thr Glu Ile Ile Arg Tyr Ser
            140                     145                     150
Phe Tyr Thr Phe Ser Leu Leu Asn His Leu Pro Tyr Leu Ile Lys
            155                     160                     165
Trp Ala Arg Tyr Thr Leu Phe Ile Val Leu Tyr Pro Met Gly Val
            170                     175                     180
Ser Gly Glu Leu Leu Thr Ile Tyr Ala Ala Leu Pro Phe Val Arg
            185                     190                     195
Gln Ala Gly Leu Tyr Ser Ile Ser Leu Pro Asn Lys Tyr Asn Phe
            200                     205                     210
Ser Phe Asp Tyr Tyr Ala Phe Leu Ile Leu Ile Met Ile Ser Tyr
            215                     220                     225
Ile Pro Ile Phe Pro Gln Leu Tyr Phe His Met Ile His Gln Arg
            230                     235                     240
Arg Lys Ile Leu Ser His Thr Glu Glu His Lys Lys Phe Glu
            245                     250
```

```
<210> 9
<211> 267
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 636866CD1

<400> 9
Met Ser Gly Cys Phe Pro Val Ser Gly Leu Arg Cys Leu Ser Arg
  1             5                    10                     15
Asp Gly Arg Met Ala Ala Gln Gly Ala Pro Arg Phe Leu Leu Thr
            20                     25                     30
Phe Asp Phe Asp Glu Thr Ile Val Asp Glu Asn Ser Asp Asp Ser
            35                     40                     45
Ile Val Arg Ala Ala Pro Gly Gln Arg Leu Pro Glu Ser Leu Arg
            50                     55                     60
Ala Thr Tyr Arg Glu Gly Phe Tyr Asn Glu Tyr Met Gln Arg Val
            65                     70                     75
Phe Lys Tyr Leu Gly Glu Gln Gly Val Arg Pro Arg Asp Leu Ser
            80                     85                     90
Ala Ile Tyr Glu Ala Ile Pro Leu Ser Pro Gly Met Ser Asp Leu
            95                    100                    105
Leu Gln Phe Val Ala Lys Gln Gly Ala Cys Phe Glu Val Ile Leu
            110                    115                    120
Ile Ser Asp Ala Asn Thr Phe Gly Val Glu Ser Ser Leu Arg Ala
            125                    130                    135
Ala Gly His His Ser Leu Phe Arg Arg Ile Leu Ser Asn Pro Ser
            140                    145                    150
Gly Pro Asp Ala Arg Gly Leu Leu Ala Leu Arg Pro Phe His Thr
            155                    160                    165
His Ser Cys Ala Arg Cys Pro Ala Asn Met Cys Lys His Lys Val
            170                    175                    180
Leu Ser Asp Tyr Leu Arg Glu Arg Ala His Asp Gly Val His Phe
            185                    190                    195
Glu Arg Leu Phe Tyr Val Gly Asp Gly Ala Asn Asp Phe Cys Pro
            200                    205                    210
Met Gly Leu Leu Ala Gly Gly Asp Val Ala Phe Pro Arg Arg Gly
            215                    220                    225
Tyr Pro Met His Arg Leu Ile Gln Glu Ala Gln Lys Ala Glu Pro
            230                    235                    240
Ser Ser Phe Arg Ala Ser Val Val Pro Trp Glu Thr Ala Ala Asp
            245                    250                    255
Val Arg Leu His Leu Gln Gln Val Leu Lys Ser Cys
            260                    265
```

```
<210> 10
<211> 329
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475576CD1

<400> 10
Met Gln Gly Gln Thr Val Val Pro Lys Asp Ser Tyr Thr Ile Ser
1               5                   10                  15
Leu Ile Gln Arg Leu Arg Gly Arg Glu Ala Ala Arg Arg Thr His
            20                  25                  30
Glu Asn Leu Leu Arg Leu Ser Ala Leu Val Arg Ser Pro Gln Thr
            35                  40                  45
Ala Ser Ile Asp Cys His Thr Trp Ser Val Ser Ser Gly Thr Asn
        50                  55                  60
Thr Ser Leu Gln Ala Ser Gly Leu Gly Arg Gln Gly Ser Cys Asp
65                  70                  75
Arg Ile Ala Ser Arg Ala Ala Ser Trp Gly Cys Thr Arg Thr Ala
            80                  85                  90
Ala Pro Gly Ile Met Gly Asn Gly Met Thr Lys Val Leu Pro Gly
            95                  100                 105
Leu Tyr Leu Gly Asn Phe Ile Asp Ala Lys Asp Leu Asp Gln Leu
        110                 115                 120
Gly Arg Asn Lys Ile Thr His Ile Ile Ser Ile His Glu Ser Pro
        125                 130                 135
Gln Pro Leu Leu Gln Asp Ile Thr Tyr Leu Arg Ile Pro Val Ala
        140                 145                 150
Asp Thr Pro Glu Val Pro Ile Lys Lys His Phe Lys Glu Cys Ile
        155                 160                 165
Asn Phe Ile His Cys Cys Arg Leu Asn Gly Gly Asn Cys Leu Val
        170                 175                 180
His Cys Phe Ala Gly Ile Ser Arg Ser Thr Thr Ile Val Thr Ala
        185                 190                 195
Tyr Val Met Thr Val Thr Gly Leu Gly Trp Arg Asp Val Leu Glu
        200                 205                 210
Ala Ile Lys Ala Thr Arg Pro Ile Ala Asn Pro Asn Pro Gly Phe
        215                 220                 225
Arg Gln Gln Leu Glu Glu Phe Gly Trp Ala Ser Ser Gln Lys Leu
        230                 235                 240
Arg Arg Gln Leu Glu Glu Arg Phe Gly Glu Ser Pro Phe Arg Asp
        245                 250                 255
Glu Glu Glu Leu Arg Ala Leu Leu Pro Leu Cys Lys Arg Cys Arg
        260                 265                 270
Gln Gly Ser Ala Thr Ser Ala Ser Ser Ala Gly Pro His Ser Ala
        275                 280                 285
Ala Ser Glu Gly Thr Leu Gln Arg Leu Val Pro Arg Thr Pro Arg
        290                 295                 300
Glu Ala His Arg Pro Leu Pro Leu Leu Ala Arg Val Lys Gln Thr
        305                 310                 315
Phe Ser Cys Leu Pro Arg Cys Leu Ser Arg Lys Gly Gly Lys
        320                 325

<210> 11
<211> 1845
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1905692CB1
```

```
<400> 11
cacaaatgaa ttatcaggag tgaacccaga ggcacgtatg aatgaaagtc ctgatccgac 60
tgacctggcg ggagtcatca ttgagctcgg ccccaatgac agtccacaga caagtgaatt 120
taaaggagca accgaggagg cacctgcgaa agaaagccca cacacaagtg aatttaaagg 180
agcagcccgg gtgtcaccta tcagtgaaag tgtgttagca cgactttcca agtttgaaga 240
tgaagatgct gaaaatgttg cttcatatga cagcaagatt aagaaaattg tgcattcaat 300
tgtatcatcc tttgcatttg gactatttgg agtttttcctg gtcttactgg atgtcactct 360
catccttgcc gacctaattt tcactgacag caaactttat attcctttgg agtatcgttc 420
tatttctcta gctattgcct tatttttttct catggatgtt cttcttcgag tatttgtaga 480
aaggagacag cagtattttt ctgacttatt taacatttta gatactgcca ttattgtgat 540
tcttctgctg gttgatgtcg tttacatttt ttttgacatt aagttgctta ggaatattcc 600
cagatggaca catttacttc gacttctacg acttattatt ctgttaagaa tttttcatct 660
gtttcatcaa aaaagacaac ttgaaaagct gataagaagg cgggtttcag aaaacaaaag 720
gcgatacaca agggatggat ttgacctaga cctcacttac gttacagaac gtattattgc 780
tatgtcattt ccatcttctg gaaggcagtc tttctataga aatccaatca aggttattcc 840
ctatagagat atgacataca tattatttat tttaggtgaa agagcttacg atcctaagca 900
cttccataat agggtcgtta gaatcatgat tgatgatcat aatgtcccca ctctacatca 960
gatggtggtt ttcaccaagg aagtaaatga gtggatggct caagatcttg aaaacatcgt 1020
agcgattcac tgtaaaggag gcacagatag aacaggaact atggtttgtg ccttccttat 1080
tgcctctgaa atatgttcaa ctgcaaagga aagcctgtat tattttggag aaaggcgaac 1140
agataaaacc cacagcgaaa aatttcaggg agtaaaaact ccttctcaga agagatatgt 1200
tgcatatttt gcacaagtga aacatctcta caactggaat ctccctccaa gacggatact 1260
ctttataaaa cacttcatta tttattcgat tcctcgttat gtacgtgatc taaaaatcca 1320
aatagaaatg gagaaaaagg ttgtcttttc cactatttca ttaggaaaat gttcggtact 1380
tgataacatt acaacagaca aaatattaat tgatgtattc gacggtccac ctctgtatga 1440
tgatgtgaaa gtgcagtttt tctcttcgaa tcttcctaca tactatgaca attgctcatt 1500
ttacttctgg ttgcacacat cttttattga aaataacagg ctttatctac caaaaaatga 1560
attggataat ctacataaac aaaaagcacg gagaatttat ccatcagatt ttgccgtgga 1620
gatactttt ggcgagaaaa tgacttccag tgatgttgta gctggatccg attaagtata 1680
gctccccctt ccccttctgg gaaagaatta tgttctttcc aaccctgcca catgttcata 1740
tatcctaaat ctatcctaaa tgttccttga agtatttatt tatgtttata tatgtttata 1800
tatgttcttc ataaatctat tacatatata tagataaaaa aaaaa        1845
```

```
<210> 12
<211> 2451
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476908CB1

<400> 12
ccggacccgg cgagccttcg gggcgcgcgt cgctggtggt ggttgaggct ctagcgataa 60
taaatgatag aggatacaat gactttgctg tctctgctgg gtcgcatcat gcgctacttc 120
ttgctgagac ccgagacgct tttcctgctg tgcatcagct tggctctatg gagttacttc 180
ttccacaccg acgaggtgaa gaccatcgtg aagtccagcc gggacgccgt gaagatggtg 240
aagagcaagg tagccgagac catgcagaac gatcgactcg gggggcttga tgtgctcgag 300
gccgagtttt ccaagacctg ggagttcaag aaccacaacg tggcggtgta ctccatccag 360
ggccggagag accacatgga ggaccgcttc gaagttctca cggatctggc caacaagacg 420
cacccgtcca tcttcgggat cttcgacggg cacggggag agactgcagc tgaatatgta 480
aaatctcgac tcccagaggc tcttaaacag catcttcagg actacgagaa agacaaagaa 540
aatagtgtat tatcttacca gaccatcctt gaacagcaga ttttgtcaat tgaccgagaa 600
atgctagaaa aattgactgt atcctatgat gaagcaggca caacgtgttt gattgctctg 660
ctatcagata aagacctcac tgtggccaac gtgggtgact cgcgcggggt cctgtgtgac 720
aaagatggga cgctattcc tttgtctcat gatcacaagc cttaccagtt gaaggaaaga 780
aagaggataa agagagcagg tggtttcatc agtttcaatg gctcctggag ggtccaggga 840
atcctggcca tgtctcggtc cctggggggat tatccgctga aaaatctcaa cgtggtcatc 900
ccagacccag acatcctgac ctttgacctg gacaagcttc agcctgagtt catgatcttg 960
gcatcagatg gtctctggga tgctttcagc aatgaagaag cagttcgatt catcaaggag 1020
cgcttggatg aacctcactt ggggccaag agcatagttt tacagtcatt ttacagaggc 1080
tgccctgaca atataaacagt catggtggtg aagttcagaa atagcagcaa aacagaagag 1140
cagtgaaccc ttcaggggtc tcagctgcct tagactaaag gactttcaac acactggtct 1200
```

```
ctttttaatttt agtgaaaagt gtgggagttg taattaggat catccaccc agacatggaa 1260
tccccctcc ctggtggtct taggtctata atcagtgacg aacagaggg gcccttggcc 1320
aatgtagtta agaaactgga aaatggtttc ttcatgtttt cccaactctt tcatccagtg 1380
tccaaaatat ataagtaaat agctgtagag tcacatatat gaagtgaata gcatatgtgt 1440
catttagtct ccctgaagat tcttttcaag atcctgttca gggtcctcca ggcatcagct 1500
gttgtgtcct ctctttgtaa cagtggacag gacagaccac ccagtgctgc aggagacagg 1560
ccactgcgtc acctgtgagt ggtcaggggc tgatgtggca acaccctctg ccaagagaca 1620
gagctgtcct gagaatgctt tgtccttctg agcccatgtt ttctgctcag tagcagcttg 1680
gaagcagatt tggaatggtt tattattttg gctgctcttg gggactgcga gaagcagaga 1740
gaatgagaga ccagtggcaa ctgcctgcac agcagagata accctcttcc cttgcttcct 1800
ttaatagtta aatagacttt gtataccacc tgaccagcct ttgtgcattt atcctaatca 1860
tgcatgaccg ttaacctttt gcttagtcct taccatatgt aataggcagc tgttaaattc 1920
accaacagat accctgattt ttcatcttac gtgaccaaga aaccacgtta ggggaaatga 1980
aaaaagcaag ccacaatacc atgattcctt ccattttcaa cagtagatga aggaaatgat 2040
actgaatgag tcacagtgtt ccctggcaag taagctgttt gcattgagaa aggagtgagc 2100
tggtgaggtt accaccctga attgagctcc agctgccagt ttttgtgttt ttccttgccc 2160
ctttccaagt ggttttcaag tgtcaggcag tgttctgaga agcagcagcc tataactgta 2220
tgtgtgttcc ttgaagccag gtgcagagtt cccagctact gcagcttggg atttggtggg 2280
aaactactgg gataagcttc tccttgacaa tggaaaggca gcagtcttca acatttggtt 2340
gcaaatctcc atccacatca gggagctttc cccaggcaaa tacaaaccgc cccgtggcct 2400
gcaggcctgc aggggaggca gcaaagggac ctggcagttg caacacagta a       2451
```

```
<210> 13
<211> 1105
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7708162CB1

<400> 13 .
agtgtgctgg aaaggtttag aatgcctctt tttcaagatg gatttccttc acaggaatgg 60
agtgctcata attcagcatt tgcagaagga ctaccgagct tactacactt ttctaaattt 120
tatgtccaat gttggagacc ccaggaatat cttttttcatt tattttccac tttgttttca 180
atttaatcag acagttggaa ccaagatgat atgggtagca gtcattgggg attggttaaa 240
tcttatattt aaatggatat tatttggtca tcgaccttac tggtgggtcc aagaaactca 300
gatttaccca aatcactcaa gtccatgcct tgaacagttc cctactacat gtgaaacagg 360
tccaggaagt ccatctggcc atgcaatggg cgcatcctgt gtctggtatg tcatggtaac 420
cgctgccctg agccacactg tctgtgggat ggataagttc tctatcactc tgcacagact 480
gacctggtca tttctttgga gtgttttttg gttgattcaa atcagtgtct gcatctccag 540
agtattcata gcaacacatt ttcctcatca agttattctt ggagtaattg gtggcatgct 600
ggtggcagag gcctttgaac acactccagg catccaaacg gccagtctgg gcacatacct 660
gaagaccaac ctctttctct tcctgtttgc agttggcttt tacctgcttc ttagggtgct 720
caacattgac ctgctgtggt ccgtgcccat agccaaaaag tggtgtgcta accccgactg 780
gatccacatt gacaccacgc cttttgctgg actcgtgaga aaccttgggg tcctctttgg 840
cttgggcttt gcaatcaact cagagatgtt cctcctgagc tgccgagggg gaaataacta 900
cacactgagc ttccggttgc tctgtgcctt gacctcattg acaatactgc agctctacca 960
tttcctccag atcccgactc acgaagagca tttatttttat gtgctgtctt tttgtaaaag 1020
tgcatccatt cccctaactg tggttgcttt cattccctac tctgttcata tgttaatgaa 1080
acaaagcgga aagaagagtc agtag                                    1105
```

```
<210> 14
<211> 1730
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473603CB1

<400> 14
atgctggagt ctgctgaaca gctgctggtg gaggacctgt acaaccgcgt cagggagaag 60
```

```
atggatgaca ccagcctcta taatacgccc tgtgtcctgg acctacagcg ggccctggtt 120
caggatcgcc aagaggcgcc ctggaatgag gtggatgagg tctggcccaa tgtcttcata 180
gctgacagga gtgtggctgt gaacaaggg aggctgaaga ggctgggaat cacccacatt 240
ctgaatgctg cgcatggcac cggcgtttac actggccccg aattctacac tggcctggag 300
atccagtacc tgggtgtaga ggtggatgac tttcctgagg tggacatttc ccagcatttc 360
cggaaggcgt actgtcatta catcattttc tcttgtgttt tcatttcagg gaaagtcctg 420
gtcagcagcg aaatgggcat cagccggtca gcagtgctgg tggtcgccta cctgatgatc 480
ttccacaaca tggccatcct ggaggctttg atgaccgtgc gtaagaagcg ggccatctac 540
cccaatgagg gcttcctgaa gcagctgcgg gagctcaatg agaagttgat ggaggagaga 600
gaagaggact atggccggga gggggatca gctgaggctg aggaggcga gggcactggg 660
agcatgctcg gggccagagt gcacgccctg acggtggaag aggaggacga cagcgccagc 720
cacctgagtg gctcctccct ggggaaggcc acccaggcct ccaagcccct caccctcata 780
gacgaggagg aggaggagaa actgtacgag cagtggaaga aggggcaggg cctcctctca 840
gacaaggtcc cccaggatgg aggtggctgg cgctcagcct cctctggcca gggtggggag 900
gagctcgagg acgaggacgt ggagaggatc atccaggagt ggcagagccg aaacgagagg 960
taccaagcag aagggtaccg gaggtgggga agggaggagg agaaggagga ggagagcgac 1020
gctggctcct cggtggggga gcggcggcgc accctgagcg agagcagcgc ctgggagagc 1080
gtgagcagcc acgacatctg ggtcctgaag cagcagctgg agctgaaccg cccggaccac 1140
ggcaggaggc gccgcgcaga ctcgatgtcc tcggagagca cctgggacgc atggaacgag 1200
aggctgctgg agattgagaa ggaggcttcc cggaggtacc acgccaagag caagagagag 1260
gaggcggcag acaggagctc agaagcaggg agcagggtgc gggaggatga tgaggacagc 1320
gtgggctctg aggccagttc cttctacaac ttctgcagca ggaacaagga caagctcact 1380
gcctggaaag atggaagatc aagagaatcc aatttggatt tcacaagaaa gacttgggag 1440
cgggagacag cagcggtgag cccggtgcag aggaggcagt aggggagaag aacccctccg 1500
acgtcagcct gacagcctac caggccctgg aagctgaaac accagaagaa ggtggggcag 1560
tgagaaccag gaggaggtgg tggagctcag caggggggag gacttggcct tggctaagaa 1620
gagacgacgg aggctggagc tgctggagag aagccggaga acctggagga gagccagtct 1680
attgcagctg ggaggcggac agtccagcgc ggggagattc cctgttggtt         1730
```

```
<210> 15
<211> 2145
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7476687CB1

<400> 15
tcctctgcag cctgccggcg cctcgaagcc cggacctgcc tccgcctctt cctccagcgg 60
ctccatcccg cctcccgtgc ctcgtcctcc cgccgcctcc gccgccgcgc ccccggtggc 120
tgcctcggcg gaccggggag ggggcccacc gcgtcggccg cccgctcggc tcggctcggc 180
ccggcccggg aggcgtgcat gcccctgctg cccgcggcgc tcaccagcag catgctctat 240
ttccagatgg tgatcatggc agggacggtg atgctggcgt actacttcga gtatacggac 300
acgttcaccg tgaacgtgca gggcttcttc tgccacgaca cgcctaccg caaaccctac 360
ccgggcccgg aggacagcag cgccgtgccc cccgtgctcc tctactcgct ggccgccggg 420
gtccccgtgc tcgtgataat agttggagaa actgctgtct tttgcctaca actagccaca 480
agggattttg aaaaccagga aaaaactat ttaactggag actgttgcta tataaacccg 540
ctggtgcgcc gaactgtccg atttcttgga atttatacat ttggactgtt tgctacagat 600
atctttgtaa atgctggaca agtagtcaca ggaaatctgg ccccacattt ccttgccctg 660
tgtaagccca attatacagc acttggatgt cagcagtata cacaattcat cagtggggaa 720
gaggcctgta ctggcaaccc agatctcatc atgagagccc gaaaaacctt tccatccaaa 780
gaagcagctc tcagtgtcta tgcagctatg tatctgacca tgtacatcac caacacaatc 840
aaagccaagg aaccagact tgctaagcca gttctatgct tgggcttaat gtgtttggca 900
tttcttactg gactcaacag agtagcagaa tatcgaaatc attggtcaga tgttatagca 960
ggctttctgg ttggaatatc tatagcagta tttctggttg tgtgcgtggt gaataatttc 1020
aaagggagac aagcagaaaa tgagcatata cacatggata atctggcaca gatgccaatg 1080
atcagcattc ctcgagtaga aagtcctttg gaaaaggtaa catctgtaca gaaccacatc 1140
actgccttcg cagaagtcac atgatatcga agcagatggt ttttcactgc attggacatc 1200
atccctttt accatccatt cataacaccc aaagtttgtt tgattgcaag tgaagtttat 1260
aaagttgttt aataattttt ataatttaa aatcaatgtc aacctatctg tttcccccac 1320
tagactgtga gctcctcgag ggcaggattg tcttttttt tctgtgtccc cagcacttac 1380
aacaaagcct agcacaaagt aggtgttcaa caaaaatgtg atgaccaaat gaaaaaaatt 1440
```

```
cttaaataaa acattcactt tagtttctca cagaatcatt gcaattatgt taaaagaaat 1500
ctctacataa atcgtatttg tgtatgaaaa ccttctattt tgggctagtt attttttttaa 1560
tcttcatata tctattcagc agtatgccat atttaatttg aagtggactt tgaaagtcat 1620
ggggggttttt attttgttat tcagcatgac attatttcca ttcgtaacat ttcagtgtgt 1680
gaaattactt tattttttaga aagtatgttc tatagtaaaa taatgtttcc acattatatt 1740
atgttatatt tcacttaaaa tactattcat actatacatt ctaagactgg tgcttctgct 1800
tttgaagggg aaaatgccaa ttttttactgt aataagtaat gtatcataat taaaaattat 1860
ttatttggac ttcttctcct gacaattgtg gcttaattca tgacttgttt ttgaatgcag 1920
gcagtattta gggtagttta aatgagtaaa ttcagcactg gtaccttatt attgagtaat 1980
ttcccatggg tagcagtgtc tcaagagtgg tcaaaagctc cactcttagg ctttttttact 2040
actaaagatt ccacataatt taaatgggaa agaactatac cctgacacat aatttaaatt 2100
atataaatgc tagaaatatg gttaatgtat tttactttgc atttg               2145
```

```
<210> 16
<211> 1352
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480440CB1

<400> 16
atggccggcc tggggttttg gggccaccct gctggacctc tcctgctgct gctgctgctg 60
gtgctgccac cccgggccct gccagaagga cccctggtgt tcgtggctct ggtattccgc 120
catggcgacc gggcccccgct ggcctcctac cccatggacc cacacaagga ggtggcctcc 180
accctgtggc cacgaggcct gggccagctg accacggagg gggtccgcca gcagctggag 240
ctgggccgct tcctgaggag ccgctacgag gccttcctga gtccggagta ccggcgggag 300
gaggtgtaca tccgcagcac ggactttgac cgcacgctgg agagtgccca ggccaacctt 360
gccgggctgt ttcccgaggc tgctccaggg agccccgagg cccgctggag gccgatcccg 420
gtgcacacgg tgcccgtggc tgaggataag ctgctgaggt tccccatgcg cagctgtccc 480
cgataccacg agctgctgcg ggaggccacc gaggccgccg agtaccagga ggccctggag 540
ggctggacgg gcttcctgag tcgcctggag aacttcacgg gactgtcgct ggttggagag 600
ccactgcgca gggcatggaa ggttctggac accctcatgt gccagcaagc ccacggtctt 660
ccactaccag cctgggcctc cccagatgtc ctgcggactc ttgcccagat ctcggctttg 720
gatattggag cccacgtggg cccacccccgg gcagcagaga aggcccagct gacaggggggg 780
atcctgctga atgctatcct tgcaaacttc tcccgggtcc agcgcctggg gctgcccctc 840
aagatggtca tgtactcagc tcatgacagc accctgctgg ccctccaggg ggccctgggc 900
ctctatgatg gacacacccc gccatatgct gcctgcctcg gctttgagtt ccggaagcac 960
ctggggaatc ccgccaaaga tggagggaat gtcaccgtct ccctcttcta ccgcaatgac 1020
tccgcccacc tgcccctgcc tctcagcctc cccgggtgcc cggcccccctg tccactaggc 1080
cgcttctacc agctgactgc cccggcccgg cctcccgccc atggggtctc ctgccatggc 1140
ccctatgagg ctgccatccc cccagctcca gtggtgcccc tgctggccgg agctgtagct 1200
gtgctggtgg cactcagctt ggggctgggc ctgctggcct ggagaccagg gtgcctgcgg 1260
gccttggggg gccccgtgtg agccagaaac cagggcttcc ctaccccccag ctgacactgg 1320
accccaacat gtatgctcag tagctgcaaa aa                             1352
```

```
<210> 17
<211> 3766
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480570CB1

<400> 17
gaaaagaaga tgaggaggag agcgacggga cgggacgcga gcgggagcgc agccgccctc 60
tcggctccgc ggcggcgcct cgcaagtccg ggaggcgagg ggggcccgag gggagacgcc 120
gtgacaactt tcgtttccct ctgagggaat tgggaggtcg gcggccccaa aagctttcag 180
tccagtgtaa agctgttgga gcgcgggagc aaaggtaaag aatgatgtaa tgcgctggct 240
gctccaaagc atcttttgtt gtggaatggt tattccagtc atctctttat gaatcaaatg 300
tgaggggctg ctttgtggac ggagtccttt gcaagagcac atcaacggga aagagaaaga 360
```

```
gacattcact tggagggctc ttgctgaaaa tgggtttaac tctcctttg ccagtcacca 420
ccagcctgac ctcatacact tttagtacaa tggagtggct gagcctttga gcacaccacc 480
attacatcat cgtggcaaat taaagaagga ggtgggaaaa gaggacttat tgttgtcatg 540
gcccatgaga tgattggaac tcaaattgtt actgagaggt tggtggctct gctgtgaaagt 600
ggaacggaaa aagtgctgct aattgatagc cggccatttg tggaatacaa tacatcccac 660
attttggaag ccattaatat caactgctcc aagcttatga agcgaaggtt gcaacaggac 720
aaagtgttaa ttacagagct catccagcat tcagcgaaac ataaggttga cattgattgc 780
agtcagaagg ttgtagttta cgatcaaagc tcccaagatg ttgcctctct ctcttcagac 840
tgttttctca ctgtacttct gggtaaactg gagaagagct tcaactctgt tcacctgctt 900
gcaggtgggt ttgctgagtt ctctcgttgt ttccctggcc tctgtgaagg aaaatccact 960
ctagtcccta cctgcatttc tcagccttgc ttacctgttg ccaacattgg gccaacccga 1020
attcttccca atctttatct tggctgccag cgagatgtcc tcaacaagga gctgatgcag 1080
cagaatggga ttggttatgt gttaaatgcc agcaatacct gtccaaagcc tgactttatc 1140
cccgagtctc atttcctgcg tgtgcctgtg aatgacagct tttgtgagaa aatttttgccg 1200
tggttggaca aatcagtaga tttcattgag aaagcaaaag cctccaatgg atgtgttcta 1260
gtgcactgtt tagctgggat ctcccgctcc gccaccatcg ctatcgccta catcatgaag 1320
aggatggaca tgtctttaga tgaagcttac agatttgtga aagaaaaaag acctactata 1380
tctccaaact tcaattttct gggccaactc ctggactatg agaagaagat taagaaccag 1440
actggagcat cagggccaaa gagcaaactc aagctgctgc acctgggagaa gccaaatgaa 1500
cctgtccctg ctgtctcaga gggtggacag aaaagcgaga cgcccctcag tccaccctgt 1560
gccgactctg ctacctcaga ggcagcagga caaaggcccg tgcatcccgc cagcgtgccc 1620
agcgtgccca gcgtgcagcc gtcgctgtta gaggacagcc cgctggtaca ggcgctcagt 1680
gggctgcacc tgtccgcaga caggctggaa gacagcaata agctcaagcg ttccttctct 1740
ctggatatca aatcagtttc atattcagcc agcatggcag catccttaca tggcttctcc 1800
tcatcagaag atgctttgga atactacaaa ccttccacta ctctggatgg gaccaacaag 1860
ctatgccagt tctcccctgt tcaggaacta tcggagcaga ctcccgaaac cagtcctgat 1920
aaggaggaag ccagcatccc caagaagctg cagaccgcca ggccttcaga cagccagagc 1980
aagcgattgc attcggtcag aaccagcagc agtggcaccg cccagaggtc ccttttatct 2040
ccactgcatc gaagtgggag cgtggaggac aattaccaca ccagcttcct tttcggcctt 2100
tccaccagcc agcagcacct cacgaagtct gctggcctgg gccttaaggg ctggcactcg 2160
gatatcttgg cccccccagac ctctacccct tccctgacca gcagctggta ttttgccaca 2220
gagtcctcac acttctactc tgcctcagcc atctacggag gcagtgccag ttactctgcc 2280
tacagctgca gccagctgcc cacttgcgga gaccaagtct attctgtgcg caggcggcag 2340
aagccaagtg acagagctga ctcgcggcgg agctggcatg aagagagccc ctttgaaaag 2400
cagtttaaac gcagaagctg ccaaatggaa tttggagaga gcatcatgtc agagaacagg 2460
tcacgggaag agctggggaa agtgggacgt cagtctagct tttcggccag catggaaatc 2520
attgaggtct cctgaagaga aagacacttg tgacttctat agacaatttt tttttcttgt 2580
tcacaaaaaa attccctgta aatctgaaat atatatatgt acatacatat atattttgg 2640
aaaatggagc tatggtgtaa aagcaacagg tggatcaacc cagttgttac tctcttaaca 2700
tctgcatttg agagatcagc taatacttct ctcaacaaaa atggaagggc agatgctagg 2760
atccccccta gacggaggaa aaccatttta ttcagtgaat tacacatcct cttgttctta 2820
aaaaagcaag tgtctttggt gttggaggac aaaatcccct accatttttca cgttgtgcta 2880
ctaagagatc tcaaatatta gtctttgtcc ggacccttcc atagtacacc ttagcgctga 2940
gactgagcca gcttgggggt caggtaggta gaccctgtta gggacagagc ctagtggtaa 3000
atccaagaga aatgatccta tccaaagctg attcacaaac ccacgctcac ctgacagccg 3060
agggacacga gcatcactct gctggacgga ccattagggg ccttgccaag gtctacctta 3120
gagcaaaccc agtacctcag acaggaaagt cggggctttg accactacca tatctggtag 3180
cccattttct aggcattgtg aataggtagg tagctagtca cactttcag accaattcaa 3240
actgtctatg cacaaaattc ccgtgggcct agatggagat aatttttttt tcttctcagc 3300
tttatgaaga gaagggaaac tgtctaggat tcagctgaac caccaggaac ctggcaacat 3360
cacgatttaa gctaaggttg ggaggctaac gagtctacct ccctctttgt aaatcaaaga 3420
attgtttaaa atgggattgt caatcctta aataaagatg aacttggttt caagccaaat 3480
gtgaatttat ttgggttggt agcagagcag cagcaccttc aaattctcag ccaaagcaga 3540
tgtttttgcc ctttctgctt cactgcatgg atacagttgg taaaatgtaa taatatggca 3600
gaattttata ggaaacttcc tagggaggta aattatggga agattaagaa aggtacaaat 3660
tgctgaggag aagcaggaaa cctgtttcct tagtggcttt tatcccctcg gcatgcgatg 3720
gggctgatgt ttctatgatt gcctcagact ttcacattta ctagta            3766
```

<210> 18
<211> 2656
<212> DNA
<213> Homo sapiens

```
<220>
<221> misc_feature
<223> Incyte ID No: 4555838CB1

<220>
<221> unsure
<222> 2532
<223> a, t, c, g, or other

<400> 18
gccgtggacc ctgtcccggc tccgccccgc ggccctggcc ccgcctgccc ctcgtccctt 60
cctccgcccc ctctcctcgc gtagtccggc ccgagccgct cgcgctagga gagcgggctt 120
cgggcacttg acatggcggc agtggcggcg actgcagcag cgaaggggaa tggggggcggc 180
ggtgggaggg ccggggccgg ggacgccagc ggcacgcgga agaagaaggg cccggggccc 240
ccggccacgg cgtacctggt catctacaat gtggtgatga cagccgggtg gctggttata 300
gcggttggtc tggtccgagc atacctggct aagggtagct accatagcct ttattattca 360
attgaaaagc ctttgaaatt ctttcaaact ggagccttat tggagatttt acattgtgct 420
ataggaattg ttccatcttc tgttgtcctg acttctttcc aggtgatgtc aagagttttt 480
ctaaatatggg cagtaacaca tagcgtcaaa gaggtacaga gtgaagacag tgtcctcctg 540
tttgttattg catggacgat cacggaaatc atccgttact ccttttatac attcagtcta 600
ttaaaccatc tgccttacct catcaaatgg gccaggtaca cactttttcat tgtgctgtac 660
ccaatgggag tgtcaggaga actgctcaca atatatgcag ctctgccctt tgtcagacaa 720
gctggcctat attccatcag tttacccaac aaatacaatt tctcttttga ctactatgca 780
ttcctgattc taataatgat ctcctacatt ccaattttc cccagttata cttccacatg 840
atacaccaga gaagaaagat cctttctcat actgaagaac acaagaaatt tgaatagttc 900
ctgctttctg cacctcccac caaaacaaac ttttcaatga tcaaaaaatg ctgcagattt 960
tttgagttcc caatacgttt catagaaaat aagtaagaac tatttttaaa atattcaaac 1020
aaaactaaaa caaaaatcca gtgtcacatg ggcctgagat tttatttttag aaaaaggttg 1080
ttacataaaa caccctggcc agttcatttc agcatgctct ttcaaccaga agttcttaat 1140
atttatgatg gcactagaaa gggatttggc attttatgtc cttctgtgtc cttcatgtat 1200
ctgatcaatg aagacctgta acactaagta cttgagagtt acagtctgaa taatgaagtc 1260
gtaccagctg aatagcccag cttgcagtat agttatgttt cagtctgcag tgtgtttagc 1320
attcccttgt caaagtgctt gactgcatgc tggaaacttt gtatttttga agcagcaaac 1380
tctgttctct ggaatgctct gaagttatgg ctgggaccta tcccctcaca tctaatgaat 1440
gaattataaa atgtatatgt ctatgaagct tcggggtagt gcctgtaatc agaaaacaac 1500
ttagaaccct tttgtttgtt tccaattgag tcattactgc ctgccactaa gaaacgtgct 1560
tgaatctaat aagtatgtgt gtaccgtaaa gaatatatct tatctggagc tcagcctcaa 1620
tcatgtctta acaaaatgac aggtctcaga aaggggggagc tcaatagctc aaaagtgaca 1680
agtcctttc acagcaccgt tctcagaaca cctctgagta acgtgtttgc cagtagctat 1740
tctcactgat gcactgatgg ccctgaagaa gcggatccag tcacatagga aaggaggctg 1800
tgttagtgaa agcacatgga aggtgttgct ttagaaaggt agtcaggaaa aacattcagg 1860
aatagattta tacaccatta ttgtttttatt tttaaatttt cattcactct tctgtttgga 1920
tacttttgct aattaacgtc ctatgttaat ttccaccaag ctataagtcc atagtcagta 1980
aaacattccc cttgggctgt catgagctaa aagcagtgtc atctccgcat gttggagcag 2040
ccaagaaata gtttggtact accgacatcg tctaatccat gtcacatcct catacaattt 2100
aattgctcaa ccatgcattt aaaactcctc aagaaaggat tggtactgca actgtaggta 2160
aactgaaaaa aaataagaaa gaaagagttg gatgaaaatg tgaaagccca agtttagatg 2220
tgcattaagt attaaatagc acagtatctt cttcatggag cctttttttcc tcccccatcc 2280
cctgcagctg cctttttttg ggggcagggt gggggttgat gttgaacttt aagagtttaa 2340
aagtttagct tattgagtag ttgtcattta aaatataatt gcgaatatca gaaaactcat 2400
actggaaaac taaattttttt tttttctctt gagacggagt ctcgctctgt tgcccaggct 2460
ggagtgcagt ggcgcgatct cggctcactg caagctccac ctcccggggtt cacgccatcc 2520
tcctgactca gnctcctgag tagctgggac tacaggtgcc tggcaacaaa cccagctaat 2580
tttttttgttt ttaagaaaaa cggggttcac cgggttaccc agatggtctg atttctgacc 2640
cttgacaccc gctaag                                                  2656

<210> 19
<211> 1292
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
```

<223> Incyte ID No: 636866CB1

<400> 19
cgctgctgca gcagccgcag cgccggccgc ggctccggct ccggctccgg ctcccgggca 60
tttaaagggg acgcggcggc tgcccggggg ggatgagggg caagtggagg ggacggctca 120
gacgcacatc atcctcagtc cctcgggact ggagggactc gtgagccgga gcccagaaat 180
ccggggggtgg ataagacacc gcgtcccctc caattcccgt aagcacccct tgctccatcc 240
tgcgccccaa tacctcagct agcccccttc cccacttctt acactccaaa ctcagccggg 300
acagacctct gctgccgccg cccccacgaa cgtgtgacga cggctggagg ccaacagagt 360
ccctacaggt ggtgctcacg gtaatgcacc gacaatgagt ggctgttttc cagtttctgg 420
cctccgctgc ctatctaggg acggcaggat ggccgcgcag ggcgcgccgc gcttcctcct 480
gaccttcgac ttcgacgaga ctatcgtgga cgaaaacagc gacgattcga tcgtgcgcgc 540
cgcgccgggc cagcggctcc cggagagcct gcgagccacc taccgcgagg gcttctacaa 600
cgagtacatg cagcgcgtct tcaagtacct gggcgagcag ggcgtgcggc cgcgggacct 660
gagcgccatc tacgaagcca tcccctttgtc gccaggcatg agcgacctgc tgcagtttgt 720
ggcaaaacag ggcgcctgct tcgaggtgat tctcatctcc gatgccaaca cctttggcgt 780
ggagagctcg ctgcgcgccg ccggccacca cagcctgttc cgccgcatcc tcagcaaccc 840
gtcggggccg gatgcgcggg gactgctggc tctgcggccg ttccacacac acagctgcgc 900
gcgctgcccc gccaacatgt gcaagcacaa ggtgctcagc gactacctgc gcgagcgggc 960
ccacgacggc gtgcacttcg agcgcctctt ctacgtgggc gacggcgcca acgacttctg 1020
ccccatgggg ctgctggcgg gcggcgacgt ggccttcccg cgccgcggct accccatgca 1080
ccgcctcatt caggaggccc agaaggccga gcccagctcg ttccgcgcca gcgtggtgcc 1140
ctgggaaacg gctgcagatg tgcgcctcca cctgcaacag gtgctgaagt cgtgctgagt 1200
ctggccgcct gcaggggggt acccgggcca acggcggagg gggcggggaa gggagattcg 1260
gcaaagacag ctttactact cccttaaaaa aa   1292

<210> 20
<211> 1325
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475576CB1

<400> 20
atgcaggggc agactgtagt tccaaaagat tcctacacta tatcccttat ccagaggctg 60
cggggccgtg aggccgcaag gagaacccat gagaaccttc ttcggctgtc tgccctagtg 120
agatccccac agacagctag catcgactgc cacacgtggt cagtttctag tggaaccaat 180
acttcgctgc aggcgtcggg cctgggccgt cagggcagct gtgaccggat cgcttcccgg 240
gcggcgagct ggggggtgcac ccggaccgcc gcccccggga tcatgggcaa tggcatgacc 300
aaggtacttc ctggactcta cctcggaaac ttcattgatg ccaaagacct ggatcagctg 360
ggccgaaata agatcacaca catcatctct atccatgagt cacccccagcc tctgctgcag 420
gatatcacct accttcgcat cccggtcgct gatacccctg aggtacccat caaaaagcac 480
ttcaaagaat gtatcaactt catccactgc tgccgcctta atggggggaa ctgccttgtg 540
cactgctttg caggcatctc tcgcagcacc acgattgtga cagcgtatgt gatgactgtg 600
acggggctag gctggcggga cgtgcttgaa gccatcaagg ccaccaggcc catcgccaac 660
cccaacccag gctttaggca gcagcttgaa gagtttggct gggccagttc ccagaagctt 720
cgccggcagc tggaggagcg cttcggcgag agccccttcc gcgacgagga ggagttgcgc 780
gcgctgctgc cgctgtgcaa gcgctgccgg cagggctccg cgacctcggc ctcctccgcc 840
gggccgcact cagcagcctc cgagggaacc ctgcagcgcg tggtgccgcg cacgccccgg 900
gaagcccacc ggccgctgcc gctgctggcg cgcgtcaagc agactttctc ttgcctcccc 960
cggtgtctgt cccgcaaggg cggcaagtga ggatgcagtc cagccgtggc tccccacttc 1020
cgactggctc ccttcggggg ctgtctgcgc cttccacgcc ccccaggacg ggcccagagg 1080
ctggggggagc cccgcggcgg cctgaaccct gcctcccgcg cccgccctgc tcgtccgcgt 1140
ctgcagtcag cgtccccaac ctgtgcgtct ctgtgtccgg gccggcctgc tgcagccacc 1200
tggtgcctta gtccttgggc tggggagggg ggcccaccct taaaggcggc gggagggggag 1260
ggagggagag tggagggttt gacgggcctg gagggtatta aagagacaca gaagaaaaaa 1320
aaaaa   1325